

1995

Molecular analyses of group 1 late embryogenic abundant (Lea) genes in soybean (*Glycine max* (L.) Merr.)

Éberson Sanches Calvo
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Genetics Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Calvo, Éberson Sanches, "Molecular analyses of group 1 late embryogenic abundant (Lea) genes in soybean (*Glycine max* (L.) Merr.)" (1995). *Retrospective Theses and Dissertations*. 10765.
<https://lib.dr.iastate.edu/rtd/10765>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Molecular analyses of group 1 late embryogenic abundant (Lea) genes in soybean
(Glycine max (L.) Merr.)

by

Éberson Sanches Calvo

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Zoology and Genetics
Major: Genetics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1995

UMI Number: 9540879

UMI Microform 9540879

Copyright 1995, by UMI Company. All rights reserved.

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI

**300 North Zeeb Road
Ann Arbor, MI 48103**

DEDICATION

This dissertation marks the end of a dream and the start of new ones. I dedicate this work to the memory of my friend Emílio Carlos Sachi who could not live long enough to see most of his dreams come true.

TABLE OF CONTENTS

GENERAL INTRODUCTION	1
Rationale	1
Literature Review	2
General Aspects of Plant Embryo Development	2
The LEA Proteins	5
Factors Affecting Gene Expression During Plant Seed Development	6
Absciscic Acid	7
Trans-Acting Elements	11
Cis-Acting Elements	13
Conclusions	19
Dissertation Organization	20
Literature Cited	21
CHAPTER 1: CLONING, MAPPING, AND ANALYSES OF EXPRESSION OF THE <i>Em</i> -LIKE GENE FAMILY IN SOYBEAN (<i>Glycine max</i> (L.) Merr.)	30
Abstract	30
Introduction	31
Materials and Methods	33
Plant Material	33
ABA and Water Deficit Treatments	33
Isolation of cDNA and Genomic Clones	34
DNA Sequencing and DNA Comparisons	35
RFLP Mapping	36

Isolation of DNA and RNA, and Hybridization Analyses	36
Reverse Transcription and PCR	38
Results	38
Screening of cDNA Library	38
Genetic Mapping of the <i>S/e</i> Gene Family	39
Sequence Analyses of Genomic Clones	40
<i>Sle</i> mRNA Accumulation in Embryos	42
<i>Sle</i> mRNA Accumulation in Response to ABA and Water Deficit	44
Discussion	45
Acknowledgments	49
References	49
 CHAPTER 2: PROMOTER ANALYSES OF A GROUP 1 LEA GENE FROM SOYBEAN (<i>Glycine max</i> (L.) Merr.) IN TRANSGENIC TOBACCO	 73
Abstract	73
Introduction	74
Experimental Procedures	76
DNA Sequence and DNA Comparisons	76
Construction of Plasmids	76
PCR Mutagenesis	77
Plant Transformations	79
GUS Assays	79
Results	80
<i>S/e</i> Promoters Contain Several Putative Regulatory Elements	80
The Role of the Distal G-Box Core Motifs	81

The Role of the Proximal ACGT Core Motif	83
The Role of the SphI Box Motif	84
The Role of the CCACAT Repeats	84
Histochemical Staining of Seeds	86
Discussion	86
References	91
 CHAPTER 3: GENERAL CONCLUSIONS	 112
Literature Cited	115
 ACKNOWLEDGMENTS	 116

GENERAL INTRODUCTION

Rationale

Five years ago the first solid molecular evidence linking the hormone abscisic acid (ABA) as a major positive regulator of late embryogenesis abundant (Lea; Dure et al., 1993) genes in wheat (Marcotte et al., 1989; Guiltinan et al., 1990) and rice (Mundy et al., 1990) was reported. These results were quite intriguing for those interested in soybean seed development mostly because the pattern of endogenous ABA accumulation in soybean embryos (Ackerson, 1984) appeared to be rather distinct from the pattern observed for rice and wheat embryos (Skriver and Mundy, 1990). Furthermore, evidence already existed showing that ABA could also be a positive regulator of transcription of soybean storage protein genes expressed during the maturation phase of seed development (Bray and Beachy, 1985). Since the molecular cloning and detailed analyses of several soybean storage protein genes were already in progress in other laboratories, we decided to clone a Lea gene from soybean as a means to understand the regulation of late embryogenesis during soybean seed development in a manner comparable to storage protein gene regulation.

Probably due to its great economic importance, soybean seed development has been subjected to an array of biochemical and physiological studies (Wilcox, 1987). Thus, the study of gene expression in soybean may ultimately establish several links between phenomena at the molecular level and those at the physiological level. Soybean is still far from being a model genetic organism such as maize or *Arabidopsis*. However, several plant genes are likely to be redundant even in the small genome of *Arabidopsis* (the Lea genes are a good example), and

will not likely be identified and cloned by standard genetic approaches. Soybean seeds, essentially comprised of embryonic material, offer substantial advantage to the application of several molecular biology approaches that rely on the purification of proteins or other molecules, library constructions, etc.

The introduction section of this Dissertation will briefly review our current knowledge of seed development in angiosperms with emphasis on the molecular biological aspects of this process.

Literature Review

General Aspects of Plant Embryo and Seed Development

In angiosperms, embryo and seed development comprise the series of events that take place between double fertilization and the germination process. Double fertilization sets the stage for a series of events characteristic to plant embryo and seed development, which includes the limited pattern formation in the embryo without formation of germ line, the accumulation of storage products in a specialized tissue, and the desiccation and acquisition of a quiescent state by the embryo and seed (Goldberg et al., 1989; Goldberg et al., 1994). Despite the noticeable variation that exists in embryo and seed development between different plant taxa (Esau, 1977), these processes can be generally divided into three phases (Gatehouse and Shirsat, 1993; West and Harada, 1993).

The first phase, referred to as early embryogenesis, starts with the double fertilization event. Fertilization of the haploid egg cell nucleus by one of the sperm nuclei initiates embryogenesis, whereas the fertilization of the diploid central cell of the embryo sac originates development of triploid endosperm which nourishes the

embryo. Embryo, endosperm and ovule-derived integuments form the seed. It is during this first phase that most of the cell divisions and all of the embryo pattern formation occur (de Jong et al., 1993). Early embryogenesis has been extensively described at the morphological level, but very little is known about the molecular nature of the events occurring during this phase, mostly due to the difficulties associated with the small size of the developing embryo and seed. Nevertheless, the development of somatic embryogenesis and *in vitro* fertilization systems, and the power of *Arabidopsis* genetics are already making significant contributions to our understanding of this process (de Jong et al., 1993; Goldberg et al., 1994).

The second phase of embryo and seed development is commonly referred to as the maturation phase and, at the molecular level, this phase of development is characterized by intensive synthesis of storage products, mainly starch, proteins, and lipids, that will be used by the germinating embryo and the resulting seedling. It has been recently found that gene expression and cell differentiation typical of the maturation phase can occur without the previous organ formation and morphogenesis that takes place during early embryogenesis (Yadegari et al., 1994). The endosperm or the embryo cotyledons will be the primary storage tissue in the seed. In the case of soybean, the endosperm is very short lived, being present only during the early embryogenesis phase (Chamberlin et al., 1994), and the two cotyledons will be the site for deposition of storage products. Therefore, a soybean seed is formed almost entirely by embryogenic tissues. During the maturation phase, embryo cells undergo growth through cell expansion (Goldberg et al., 1989). Although cell divisions have almost come to an end at the maturation phase, DNA replication without cell division may occur in the cotyledons (Marks and Davies, 1969) or in the endosperm (Lopes and Larkins, 1993) of some plant species. Work done in soybean embryos (Fischer and Goldberg, 1982; Walling et

al., 1986) and maize endosperm (Lopes and Larkins, 1993) suggests that this genomic amplification is not selective, and therefore it is still questionable whether or not endoreplication is a mechanism used by the cell to accumulate large amounts of storage products.

The third phase of embryo and seed development, referred to as late embryogenesis, is marked by a dramatic reduction in water content in the cell. Transcription is also reduced, particularly from the highly active genes involved in the synthesis of storage proteins and lipids during maturation phase (Gatehouse and Shirsat, 1993). At the time that overall transcription is reduced, a new set of proteins begins to be synthesized, the so called "late embryogenic abundant" (LEA) proteins (Galau et al., 1986; Dure, 1993a). At the end of late embryogenesis the seed is mature and dry. Surprisingly, dry seed nuclei are transcriptionally active, having about 8% of maximal transcriptional rates found during early embryogenesis (Comai and Harada, 1990). More interestingly, only those genes that are transcriptionally active during late embryogenesis are capable of being transcribed in the dry seed (Comai and Harada, 1990).

Information has been obtained at the molecular level during the past fifteen years showing that embryo and seed development is a complex process, with as many as 20,000 different mRNA present at any stage (Goldberg et al., 1989). Further, it has been shown that seed development is temporally regulated. Measurements of abundance of 47 embryo-specific mRNA in cotton embryos revealed that the pattern of accumulation of these mRNA is temporally modular in this species (Hughes and Galau, 1989). In fact, the combination of five of these modules or programs could explain the kinetics of accumulation of the 12 distinguished, coordinately expressed mRNA classes identified among the 47 mRNA analyzed. These five programs are intimately associated with the

developmental stages of the cotton embryo (Hughes and Galau, 1989; Hughes and Galau, 1991]) and were named for the seed developmental stage in which they are being predominantly expressed. Thus, they are called the Cotyledonary (early embryogenesis), Maturation (storage of reserves), Post-abscission (late embryogenesis), and Germination programs. A fifth program, named ABA, was found to be correlated with the transient high levels of ABA during the maturation stage (Hughes and Galau, 1989). Evidence exists showing that similar programs also occur in other dicotyledonous species such as soybean, tobacco, oilseed rape, and *Arabidopsis* (Jakobsen et al., 1994; Parcy et al., 1994). The presence of these developmental programs in monocotyledonous species awaits further investigation.

Finally, evidence has accumulated showing that seed development is also spatially regulated. The accumulation of storage protein genes in soybean developing seeds is a classical example (Goldberg et al., 1989; Perez-Grau and Goldberg, 1989). Nevertheless, the precise mechanisms by which gene expression during seed development is regulated remains largely unknown.

The LEA Proteins

Among the developmental programs described above, the late embryogenesis program bears particular interest to this dissertation. Several Lea genes have been isolated and sequenced in different plant species (Dure, 1993a). Their polypeptides can be grouped into at least three structurally distinct groups based on amino acid sequence homology and occurrence of amino acid motifs (Dure et al., 1989). Group 1 polypeptides are small, extremely hydrophilic and highly conserved across all species. Examples of this class are the proteins

encoded by the *Em* gene from wheat (Litts et al., 1987), D19 from cotton (Baker et al., 1988), *Emb-1* from carrot (Ulrich et al., 1990), and *AtEm1* and *AtEm2* from *Arabidopsis* (Finkelstein, 1993; Gaubier et al., 1993; Calvo et al., 1994). Group 2 and 3 polypeptides are also hydrophilic but homology is limited to tracts of amino acids. Examples of genes encoding Group 2 and 3 proteins are the rice *Rab21* (Mundy and Chua, 1988), the cotton D11 (Baker et al., 1988) for Group 2 and the cotton *D7* and *D29* (Baker et al., 1988) for Group 3.

It has been hypothesized that the LEA proteins are involved in protecting plant cells from the harms of desiccation (Dure et al., 1989). Support for this hypothesis come mostly from circumstantial evidence, such as their abundance in the cell, the time of LEA protein appearance during development, and their hydrophilic nature (Dure, 1993a). Experimental evidence is limited to biophysical measurements of hydrodynamic properties of the wheat *Em* protein (McCubbin et al., 1985). Computer modeling of Group 3 proteins has predicted a tertiary structure that is compatible with the hypothesis that such proteins would be involved in ion sequestration in the desiccated cell (Dure, 1993b). Whether or not such a structure occurs *in vivo* remains to be determined. The recent finding of *Lea* genes in cyanobacteria, an organism more easily studied genetically than higher plants, will certainly contribute to our understanding of the functions of LEA proteins in late embryo development (Curry and Walker-Simmons, 1993).

Factors Affecting Gene Expression During Seed Development

Seed development has been intensively described at the morphological and physiological levels (Esau, 1977). Several studies indicate that this process is coordinately regulated in response to both developmental and environmental cues

(Goldberg et al., 1989; Goldberg et al., 1994). However, our knowledge of the molecular events at the level of gene expression, involved in the regulation of seed development, is still very limited. Rather than list all the physiological factors that were found to affect gene expression during seed development, we will present our current knowledge on the molecular events for which more than simple "stimuli vs. response" types of relationships have been established.

It is generally accepted that gene expression during seed development is accomplished primarily through regulation of transcription (Walling et al., 1986) although several studies have also suggested that regulation at the post-transcriptional level also plays a significant role (Chappell and Chrispeels, 1986; Goldberg et al., 1989; Jofuku et al., 1989). Since our knowledge of post-transcriptional regulation is at most minimal, emphasis will be given to transcriptional regulation at the maturation and late embryogenesis phases of seed development.

Abscisic Acid

Of all factors that possibly modulate seed development, the hormone abscisic acid (ABA) is by far the most studied (Quatrano, 1986; Skriver and Mundy, 1990; Quatrano et al., 1993). Particularly, it has been implicated as a positive regulator of the expression of seed storage protein genes from soybean (Bray and Beachy, 1985) and rape (Finkelstein et al., 1985) as well as of several Lea genes (Skriver and Mundy, 1990; Hughes and Galau, 1991) in *in vitro* cultured embryos. However, the interpretation of such experiments has been questioned because *in vitro* cultured embryos may not correctly mimic the conditions *in vivo* (Galau et al., 1991; Hughes and Galau, 1991).

More compelling evidence that ABA can affect both storage protein and Lea gene expression *in planta* is provided by studies with mutants. Two classes of ABA mutants have been described in maize and in *Arabidopsis thaliana* (Robertson, 1955; Koornneef et al., 1982; Koornneef et al., 1984) where they are referred to as viviparous (VP mutants) because of their expression of the precocious embryo germination phenotype. The first class is formed by mutants which are defective in ABA synthesis (e.g. *vp2*, *vp5*, *vp7*, *vp8*, *vp9*, in maize and *aba1* in *Arabidopsis*) and have lowered levels of endogenous ABA in the embryo (Robichaud et al., 1980; Karssen et al., 1983). The second class comprise the mutants in which embryos have normal levels of ABA but are impaired in ABA recognition (e.g. maize *vp1* and *Arabidopsis abi3*) (Koornneef et al., 1984; McCarty et al., 1989). The *vp1* and *abi3* mutants are of particular interest because their phenotypes are restricted to seed development and include ABA-insensitivity, desiccation intolerance, and reduced accumulation of several proteins including storage proteins and Lea proteins (Koornneef et al., 1989; Kriz et al., 1990; McCarty et al., 1991; Finkelstein, 1993).

Studies with *aba* and *abi* single and double mutants suggested that the ABI3 protein, besides inhibiting precocious germination, also has a direct role in mediating ABA-regulated gene expression during seed development, possibly by lowering threshold requirements for ABA action (Koornneef et al., 1989). Although the importance of *Vp1* or *Abi3* in dormancy acquisition or maintenance through ABA action has been well recognized, the failure of *vp1* or *abi3* mutants in expressing the maturation or late embryogenesis programs have also been interpreted as being caused by the incapacity of such embryos to enter the appropriate developmental stage (Galau et al., 1991). Since *vp1* or *abi3* embryos are early germinating embryos (Nambara et al., 1995) there would be no need for

the expression of maturation and late embryogenesis program (Galau et al., 1991). According to these authors ABA plays no direct role in *Lea* gene expression during seed development in dicot species.

A direct role of ABA in gene regulation would account for the exogenously applied ABA-responsiveness of several genes expressed during seed development (Bray and Beachy, 1985; Finkelstein et al., 1985; Skriver and Mundy, 1990; Hughes and Galau, 1991). It would also explain results of Naito (1994) showing that the accumulation of the β -subunit from the β -conglycinin storage protein is significantly reduced in *Arabidopsis aba1-1* auxotrophic mutants, as predicted by experiments with *in vitro* cultured embryos (Bray and Beachy, 1985).

Results obtained from the cloning and analyses of the maize *Vp1* gene and other *Vp1* alleles also suggest that normal levels of VP1 protein and endogenous ABA are required for correct expression of the "maturation program" in maize embryos (McCarty et al., 1991; McCarty and Carson, 1991) (which in their definition also includes late embryogenesis). Transient expression experiments with overexpression of *Vp1* in the presence of the *Em*-GUS reporter gene showed an *Em* promoter-specific synergistic interaction between *Vp1* and the hormone ABA in maize protoplasts (McCarty et al., 1991). Since *Vp1* action is seed specific the authors propose that the VP1 protein "potentiates an ABA response in seed tissues destined to undergo maturation" (McCarty et al., 1991). Indeed, analyses of *Arabidopsis* transgenic plants expressing *Ab3* in vegetative tissue showed that they were able to express several seed specific mRNAs in vegetative tissues in an ABA-dependent manner (Parcy et al., 1994).

Nevertheless, ectopic expression of *Ab3* was not capable of inducing the expression of all seed specific mRNA analyzed after ABA treatment, including members of the same gene family such as the Group 1 *Lea* genes *AtEm1* and

AtEm2 (Parcy et al., 1994). Analyses of multiple mRNA markers showed that mRNAs belonging to the same temporal class were differentially affected by the *ab3-4* mutant (Parcy et al., 1994). Moreover, characterization of ABA-deficient mutants showed that the accumulation of these mRNA is not correlated with seed ABA content. Therefore, although endogenous ABA content participates in the expression of several seed-specific genes, other developmental factors must be involved for the correct expression of maturation and post-abscission programs.

The absence of a good correlation between endogenous ABA content and *Lea* gene expression was also observed in *in vitro* cultured embryos (Galau et al., 1991; Hughes and Galau, 1991; Paiva and Kriz, 1994). Although the patterns of *Lea* gene expression in soybean have not been determined, it would be particularly difficult to envision a direct role of endogenous ABA levels in *Lea* gene expression since the peak of endogenous ABA concentration is at least 30 days before the expected peak for *Lea* gene expression (Ackerson, 1984; Ackerson, 1984).

Results obtained by Williamson and Scandalios (1992) have shown that although the maize ABA deficient *vp5* mutants have reduced levels of Catalase1 (*Cat1*) mRNA, *vp1* mutants as well as their wild type siblings respond to exogenous applied ABA. Similar results were observed for the maize *Rab28* *Lea* gene (Pla et al., 1991). Therefore, a *Vp1* independent ABA pathway does exist in maize embryos. Whether these results can be extended to other plant species remains to be determined.

Trans-acting elements

Several trans-acting factors have been implicated in regulation of seed development. The role of VP1 as a transcription factor (McCarty and Carson, 1991) was mentioned in the preceding section. The results obtained by Parcy et al. (1994) also suggest that the ABI3 protein (Giraudat et al., 1992), rather than being confined to the ABA-response pathway, has a broader role in embryogenesis than the one initially proposed by McCarty et al. (1991) for the VP1 protein, the ABI3 putative counterpart in maize. Attempts to show *in vitro* binding of VP1 to postulated target DNA sequence (discussed in the next section) have been unsuccessful (Hattori et al., 1992). Taken together these results may suggest that VP1 and ABI3 interact with several different proteins in embryonic tissues. The nature of these factors and the mechanism(s) of VP1 or ABI3 action have yet to be determined. Two new ABA-insensitive *Arabidopsis* mutants, *abi4* and *abi5*, have been recently found which identify two new loci supposedly located in the same transduction pathway as *ab3* (Finkelstein, 1994).

Two other mutants, *fus3* and *lec1* that appear to function as regulatory proteins during embryogenesis in *Arabidopsis* have been described (Bäumlein et al., 1994; Keith et al., 1994; West et al., 1994). Although these mutants produce some of the same phenotypes observed in the *ab3* embryos, they most likely participate in two new distinct regulatory pathways. While both storage protein and Lea gene expression is affected in *ab3* embryos (Parcy et al., 1994), *lec1* and *fus3* embryos appear to be defective only in storage protein gene expression (Bäumlein et al., 1994). Each of the three mutations appear to inactivate different genes having similar temporal patterns of expression, therefore suggesting the existence

of different transduction pathways acting upon the same program, i.e, the maturation program (West et al., 1994).

Genetic analyses have also enabled the isolation of the Opaque-2 (*O2*) locus in maize (Hartings et al., 1989; Schmidt et al., 1990). The protein encoded by *O2* is a member of the "leucine zipper" (bZIP) family of transcription factors and both genetic (Schmidt, 1993) and molecular biology (Schmidt et al., 1992; Ueda et al., 1992) evidence identify the *O2* protein as a primary regulator of zein storage protein expression in maize endosperm. A second bZIP protein, OHP1, that is capable of interaction *in vitro* with the *O2* protein has also been isolated from maize through homology probing (Pysh et al., 1993). The significance of this interaction *in vivo* is yet unknown. Sequence homology has also enabled the isolation of a rice gene, *Rita-1*, encoding a protein very similar to *O2* which was shown to be involved in regulating gene expression in rice seeds (Izawa et al., 1994). Interestingly, contrary to *O2*, *Rita-1* expression is not confined to the endosperm, but is also found in other non-seed tissues (Izawa et al., 1994). The occurrence of similar proteins in dicot species have not been demonstrated.

DNA- protein interactions have provided the means for the isolation of several putative bZIP transcription factors in plants (Foster et al., 1994). Many of these proteins are able to bind sequence motifs present on the promoter of a variety of seed-specific genes. Among these, the wheat *Embp-1* bears particular interest because it was isolated based on its ability to bind multimers of the ABA-responsive element (ABRE; discussed in the next section) of the wheat *Em* promoter (Guiltinan et al., 1990). However, it is still unknown whether or not *Embp-1* or any of these proteins act as a transcription activator during ABA response or normal embryo development.

With the exception of PKABA1 from wheat, all of the *trans*-acting factors cloned so far appear to act as transcriptional activators. Therefore, our knowledge of the steps occurring upstream in the signal transduction pathway(s) governing seed development is minimal. PKABA1 most likely encodes a serine/threonine protein kinase which accumulates in mature wheat embryos and is up-regulated by ABA and water stress (Anderberg and Walker-Simmons, 1992). Again, a direct involvement of PKAB-1 in ABA response or normal seed development has not yet been shown.

Cis-acting elements

The characterization of DNA sequence motifs also has proven to be a powerful tool for the isolation of regulatory proteins in signal transduction pathways in plants. Besides the transcription activators mentioned on the previous section, other types of *trans*-acting factors have also been isolated on the basis of their ability to interact with a given DNA sequence motif (Vetten et al., 1992).

One of the first seed-specific promoters to be analyzed in plants was the promoter from the soybean gene encoding the α' subunit of the β -conglycinin storage protein (Chen et al., 1986). A seed-specific enhancement activity was found in a 170 bp DNA fragment of this promoter. Initial studies suggested that the 6 bp repeat (A^{A/G}CCCCA) present in five copies in the 170 bp fragment plays a major role in the regulation of the β -conglycinin and possibly other seed specific genes (Chen et al., 1986; Chen et al., 1988; Goldberg et al., 1989). This repeat was found to be the core for specific binding of embryo nuclear protein extracts *in vitro* (Allen et al., 1989; Lessard et al., 1991). However, attempts to correlate this protein binding ability with enhancer activity in transgenic tobacco plants have

been unsuccessful (Lessard et al., 1993; Fujiwara and Beachy, 1994), clearly demonstrating that *in vitro* protein binding does not necessarily correlate with gene expression *in planta*.

Analyses of the wheat *Em* Group 1 *Lea* gene resulted in the identification of the first DNA sequence capable of conferring ABA response to a minimal promoter, the so called ABA-responsive element (ABRE), with the sequence CACGTG (Marcotte et al., 1989; Guiltinan et al., 1990; Quatrano et al., 1993). This same sequence was also shown to be responsible for the ABA response of other *Lea* genes (Mundy et al., 1990; Pla et al., 1993). Variants of these sequences have also have been implicated in ABA response. Although some of these sequences share a common ACGT core with the typical ABRE (Thomas, 1993), others are considerably different (Lam and Chua, 1991; Shen et al., 1993; Thomas, 1993). It is unclear if these sequences potentiate the ABA-response through the same transduction pathway(s). Although the ABRE and ABRE-like motifs are able to confer ABA responsiveness to a minimal promoter, it is likely that in their promoter activity *in vivo*, they do so by interacting with other motifs. Support for this hypothesis came from analyses of a barley ABA-responsive gene where the ABA-responsive complex is formed by an ABRE-like sequence and a 'Coupling Element' (CE-1;TGCCACCGG; Shen and Ho, 1995).

In all cases investigated the ABRE and its variants also confer high levels of expression in tobacco seeds (Marcotte et al., 1989; Lam and Chua, 1991; Thomas, 1993). It remains to be determined how related are these responses. Two other DNA sequence motifs have been implicated in seed-specific expression; the GTACGTGGC (Salinas et al., 1992) and the TGAC-like (TGAGTCATCA) motifs (dePater et al., 1993). The latter most likely bind a bZIP protein with binding specificity similar to the O2 in maize (de Pater et al., 1994).

Two potential binding sites have been identified for the maize O2 protein in two different promoters; the GATGAPyPuTGpu and TCCACGTTAGA motifs in the *b*-32 and α -zein promoters, respectively (Lohmer et al., 1991; Schmidt et al., 1992). Although both sites do contain the ACGT core motif usually associated with the binding of bZIP proteins (Foster et al., 1994) such as O2, it has been shown that O2 may bind to and transactivate promoters without the ACGT core (de Pater et al., 1994; Yunes et al., 1994).

The prolamin gene promoters of monocot species have a highly conserved motif usually referred to as the endosperm box (TGTAAG), which has been implicated in endosperm-specific expression in tobacco (Colot et al., 1987). Transient expression assays in endosperm-derived maize protoplasts and *in vivo* footprint analysis in wheat also support this hypothesis (Quayle and Feix, 1992; Hammond-Kosack et al., 1993).

The DNA sequence motif CANNTG, which in animal cells (Lüscher and Eisenman, 1990) acts as the binding site for the helix-loop-helix (bHLH) family of transcription factors, was also shown to be a regulatory motif in the promoter of a β -phaseolin storage protein gene (Kawagoe and Murai, 1992; Kawagoe et al., 1994). Characterization of similar sequences of the promoters of other seed-specific genes and cloning of genes encoding its binding proteins will undoubtedly be a major area of research of seed development during the next couple of years.

Sequence comparison of the promoter regions of several seed-specific genes have revealed a highly conserved motif known as the RY repeat or SphI box, comprised of the sequence CATGCATG (Dickinson et al., 1988; Hattori et al., 1992). In dicot species the SphI box has been shown to act as an enhancer in the promoter of the *Vicia faba* legumin gene (*LeB4*; Bäumlein et al., 1992), the soybean glycinin, *Gy2*, (Lelievre et al., 1992), and of the β -conglycinin α' subunit

(Chamberland et al., 1992) genes. Moreover, in the legumin promoter the destruction of the SphI box results in low levels of expression in leaves (Bäumlein et al., 1992), suggesting a role in determining seed specificity. In the β -conglycinin α' subunit promoter, the deletion of the SphI boxes resulted in activity in the leaves of transgenic tobacco plants only when the minimal core promoter was the CaMV 35S (-90 bp) and not with the α' promoter (Fujiwara and Beachy, 1994). Therefore, seed specificity in the soybean α' subunit gene is attained not solely by the presence of the SphI box, but rather through interactions of this motif with other yet unidentified sequences in the α' core promoter. Evidence also exists showing that similar interactions may be necessary for the enhancement effect of the SphI box (Bäumlein et al., 1992). In at least one case, the promoter of the *Usp* (unknown storage protein) gene in *Vicia faba*, interaction with other sequences in the promoter may explain the surprising negative effect played by the SphI box in gene expression (Fiedler et al., 1993). Interestingly, the USP gene is expressed earlier in seed development when compared to the *LeB4* legumin gene (Bäumlein et al., 1992), and the authors speculate that the same SphI box binding factor(s) are involved in down-regulation of USP and up-regulation of *LeB4* (Fiedler et al., 1993).

The SphI box has also been shown to be involved in expression of a maize seed-specific non-storage protein gene through an ABA and Vp1-dependent pathway. Hattori et al. (1992) showed that a mutation in the SphI box completely disrupts the transactivation of the *C1* promoter by both Vp1 and ABA in electroporated maize protoplasts. Moreover, they were able to show that a deletion of a 5 bp tandem repeated sequence (GTGTC) flanking the SphI box abolishes the ABA response but does not affect the Vp1 response. This deletion mimics a natural mutant allele of *C1* (*c1-p*), in which kernels reach maturity without accumulation of

anthocyanin pigments in the aleurone (McCarty and Carson, 1991), thus linking anthocyanin synthesis to the high levels of ABA during kernel late stages of development. This result also suggests that the SphI box is interacting with nearby surrounding DNA sequences to produce its developmental response.

Whether all Vp1/ABA mediated transcription activation relies on the presence of the SphI box is still unknown. Expression of the two *Em* homologous *Lea* genes in *Arabidopsis*, *AtEm1* and *AtEm6*, was found to be severely reduced in *abi3* mutant embryos (Gaubier et al., 1993). However, no SphI box was detected in the promoter region of either gene. At present it is unknown whether this discrepancy reflects differences between the VP1 and ABI3 proteins themselves or if it suggests that sequences other than the SphI box can be target sites for VP1 or ABI3 action. Several potential ABA responsive sequences (CACGTG core) were found in the promoter of the *Arabidopsis AtEm1* and *AtEm6* genes (Finkelstein, 1993; Gaubier et al., 1993; Calvo et al., 1994). Nevertheless, similar sequences located in the promoter of the *C1* gene do not appear to substantially affect the ABA response in maize protoplasts (Hattori et al., 1992). The sequence motifs required for VP1-mediated transactivation on ABA-independent genes are also unknown.

Sequences common to almost all plant genes, such as the A/T-rich domains, are almost invariably present in the promoters of seed-expressed genes (Thomas, 1993) where they appear to act as general enhancers of gene expression (Bustos et al., 1989; Jordano et al., 1989). Evidence has been recently presented showing that this enhancement is mediated through the binding of plant nuclear matrix proteins (van der Geest et al., 1994).

So far we have reviewed several sequence motifs having positive quantitative effects in gene expression during seed development. There are considerably fewer reports of the role played by *cis*-acting elements in the temporal

or spatial pattern of gene expression. This likely results not only from technical constraints, but also from conceptual difficulties in separating these effects from their quantitative counterparts.

Nevertheless, promoter regions that do not contribute quantitatively to gene expression but rather are involved in temporal control have been reported for some seed-specific genes, although the specific sequence responsible for such effects remains to be determined (Bustos et al., 1991; Burow et al., 1992; Kawagoe and Murai, 1992; Fiedler et al., 1993). Despite these results, it is likely that most of the temporal and spatial patterns of expression are determined by interactions with sequences having remarkable quantitative effect, as predicted by the combinatorial model of promoter function (Benfey and Chua, 1990).

Similarly, analyses of differentially expressed 2S albumin genes in *Arabidopsis* identified a 67 bp DNA fragment necessary for 2S albumin gene-specific expression in cotyledonary cells of palisade parenchyma and adaxial epidermis (Conceição and Krebbers, 1994). Further identification of the precise DNA sequence within the 67 bp fragment involved in this response has yet to be reported. However, putative candidate sequences including a SphI box-like sequence (CATGCA) (Conceição and Krebbers, 1994), clearly illustrate the functional overlapping of sequence motifs involved in determining quantitative and spatial expression.

Finally, it is important to mention that although the vast majority of DNA sequence motifs characterized so far have a positive role in gene regulation during seed development, DNA motifs which confer a negative effect do exist (Bustos et al., 1991; Burow et al., 1992; Kawagoe and Murai, 1992; Fiedler et al., 1993; Lessard et al., 1993). However, with the exception of the unusual negative effect of

the SphI box in the USP promoter (Fiedler et al., 1993), the precise DNA sequences of these motifs are still unknown.

Conclusions

Although our knowledge of gene expression during seed development has grown substantially in the past ten years, several basic questions remain unresolved. Among them is how embryo-specific gene expression is attained?, and how the different temporal patterns of gene expression are established in the seed?.

It becomes clear from the literature review that most of what we know about gene expression during seed development is centered on the effect of the hormone ABA. While it is incontestable that ABA- mediated pathways have profound effects on seed development, it is also evident that other factors are also involved. The initial question that sparked this research project was whether endogenous ABA is a major factor in soybean late embryo development as had been proposed at the time for other plant species. We reasoned that the cloning and analyses of Lea genes could shed some light onto this question. Moreover, it would also provide us with the means to answer other basic questions concerning gene expression during seed development in an agronomically important crop.

The fact that a simple temporal pattern of expression, such as that seen in the maturation or the late embryogenesis programs, may be determined by several different, independent transduction pathways, gives us a glimpse of the complexity behind regulation of gene expression during seed development. Therefore, a complete understanding of this process only will be possible through the analyses of several seed-specific genes in different plant species.

Dissertation Organization

This dissertation is written in an alternate format with a general introduction followed by three chapters. The first two chapters each constitutes an independent paper that have been submitted (Chapter 1) or will be submitted (Chapter 2) for publication in a refereed journal. Figures and tables are placed at the end of each paper right after the References section. Each chapter reflects solely my work under guidance from Dr. Randy C. Shoemaker. A general conclusion is presented in Chapter 3. References cited within the General Introduction follow the General Introduction section.

This dissertation not only provides the groundwork (Chapter 1) for the study of *Lea* gene expression during soybean seed development, but also begins to analyze the DNA sequence requirements for correct *Lea* gene regulation in soybean seeds (Chapter 2).

In Chapter 1 we show that the *S/e* family in soybean has five members which map to at least four independent linkage groups. These genes could be divided into two classes having distinct temporal patterns of expression during soybean seed development. We also address the expression of the *S/e* genes upon exogenous ABA treatment and water stress. This chapter has been submitted for publication in *Plant Molecular Biology*. Dr. Eve Wurtele is a co-author of this paper. She provided us with the heterologous EMB-1 cDNA probe and many valuable discussions of the results.

In Chapter 2 the promoters of four *S/e* genes are sequenced and specific mutations in the *Sle1* promoter are analyzed in transgenic tobacco plants. These experiments allowed us to identify the *SphI* box (Dickinson et al., 1988; Hattori et al., 1992) and its flanking CCACAT repeats as the major site for high quantitative

regulation of the *S/e1* promoter. This chapter will be submitted for publication in *The Plant Journal*. Dr. Eve Wurtele is also a co-author of this paper.

Literature Cited

- Ackerson R.C. (1984) Regulation of soybean embryogenesis by abscisic acid. *J. Exp. Bot.* 35: 403-413.
- Allen R.D., Bernier F., Lessard P.A., Beachy R.N. (1989) Nuclear factors interact with a soybean β -conglycinin enhancer. *Plant Cell* 1: 623-631.
- Anderberg R.J., Walker-Simmons M.K. (1992) Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. *Proc. Natl. Acad. Sci. USA* 89: 10183-10187.
- Baker J., Steele C., III L.D. (1988) Sequence and characterization of 6 *Lea* proteins and their genes from cotton. *Plant Mol. Biol.* 11: 277-291.
- Bäumlein H., Miséra S., LuerBen H., Kolle K., Horstmann C., Wobus U., Muller A. (1994) The *FUS3* gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *Plant J.* 6: 379-387.
- Bäumlein H., Nagy I., Villarroel R., Inze D., Wobus U. (1992) Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2: 233-239.
- Benfey P.N., Chua N. (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* 250: 959-966.
- Bray E.A., Beachy R.N. (1985) Regulation by ABA of β -conglycinin expression in cultured developing soybean cotyledons. *Plant Physiol.* 79: 746-750.
- Burow M.D., Sen P., Chlan C.A., Murai N. (1992) Developmental control of the β -phaseolin gene requires positive, negative, and temporal seed-specific transcriptional regulatory elements and a negative element for stem and root expression. *Plant J.* 2: 537-548.
- Bustos M.M., Begum D., Kalkan F.A., Battraw M.J., Hall T.C. (1991) Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. *EMBO J.* 10: 1469-1479.

- Bustos M.M., Guiltinan M.J., Jordano J., Begun D., Kalkan F.A., Hall T.C. (1989) Regulation of β -glucoronidase expression in transgenic tobacco plants by an A/T rich, *cis*-acting sequence found upstream of a French bean β -phaseolin gene. *Plant Cell* 1: 839-853.
- Calvo E.S., Rodermeier S., Shoemaker R.S. (1994) A third highly conserved group 1 Lea gene from *Arabidopsis thaliana*. *Plant Physiol.* 106: 787-788
- Chamberland S., Daigle N., Bernier R. (1992) The legumin boxes and the 3' part of a soybean b-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. *Plant. Mol. Biol.* 19: 937-949.
- Chamberlin M.A., Horner T.H., Palmer R.G. (1994) Early endosperm, embryo and ovule development in *Glycine max* (L.) Merr. *Int. J. Plant Sci.* 155: 421-436.
- Chappell J., Chrispeels M.J. (1986) Transcriptional and posttranscriptional control of phaseolin and phytohemagglutinin gene expression in developing cotyledons of *Phaseolus vulgaris*. *Plant Physiol.* 81: 50-54.
- Chen Z.-L., Schuler M.A., Beachy R.N. (1986) Functional analysis of regulatory elements in a plant embryo-specific gene. *Proc. Natl. Acad. Sci. USA* 83: 8560-8564.
- Chen Z.-L., Pan N.-S., Beachy R.N. (1988) A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. *The EMBO J.* 7: 297-302.
- Colot V., Robert L.S., Kavanagh T.A., Bevan M.W., Thompson R.D. (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. *EMBO J.* 6: 3559-3564.
- Comai L., Harada J.J. (1990) Transcriptional activities in dry seed nuclei indicate the timing of the transition from embryogeny to germination. *Proc. Natl. Acad. Sci. USA* 87: 2671-2674.
- Conceição A.S., Krebbers E. (1994) A cotyledon regulatory region is responsible for the different spatial expression patterns of *Arabidopsis* 2S albumin genes. *Plant J.* 5: 493-505.
- Curry J., Walker-Simmons M.K. (1993) Sequence analysis of wheat cDNAs for abscisic acid-responsive genes expressed in wheat seedlings and the cyanobacterium, *Anabaena*. In: Close TJ, Bray EA (ed.) *Plant response to cellular dehydration during environmental stress*, 128-136. American Society of Plant Physiologists, University of California, Riverside .

- de Jong A.J., Schmidt E.D.L., Vries S.C.d. (1993) Early events in higher-plant embryogenesis. *Plant Mol. Biol.* 22: 367-377.
- de Pater S., Katagiri F., Kijne J., Chua N. (1994) bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. *The Plant J.* 6: 133-140.
- de Pater S., Pham K., Chua N., Memelink J., Kijne J. (1993) A 22-bp of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. *Plant Cell* 5: 877-886.
- Dickinson C.D., Evans R.P., Nielsen N.C. (1988) RY repeats are conserved in the 5' flanking region of legume seed protein genes. *Nucl. Acids Res.* 16: 371.
- Dure III.L. (1993a) The lea proteins of higher plants In: Ed. Verma DPS *Control of Plant Gene Expression*, 325-335. CRC Press, Inc., Boca Raton, FL.
- Dure III.L. (1993b) A repeating 11-mer amino acid motif and plant desiccation. *Plant J.* 3: 363-369.
- Dure III.L., Crouch M., Harada J., Ho T.D., Mundy J., Quatrano R., Thomas T., Sung Z.R. (1989) Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol. Biol.* 12: 475-486.
- Esau K. (1977) *Anatomy of Seed Plants*. John Wiley, New York .
- Fiedler U., Filistein R., Wobus U., Baumlein H. (1993) A complex ensemble of *cis*-regulatory elements controls the expression of a *Vicia faba* non-storage seed protein gene. *Plant Mol. Biol.* 22: 669-679.
- Finkelstein R.R. (1993) Absciscic acid-insensitive mutations provide evidence for stage-specific signal pathways regulating expression of an *Arabidopsis* late embryogenesis-abundant (*lea*) gene. *Mol Gen Genet* 238: 401-408.
- Finkelstein R.R. (1994) Mutations at two new *Arabipopsis* ABA response loci are similar to the *abi3* mutations. *Plant Cell* 5: 765-771.
- Finkelstein R.R., Tenbarger K.M., Shumway J.E., Crouch M.L. (1985) Role of ABA in maturation of rapeseed embryos. *Plant Physiol.* 78: 630-636.
- Fischer R.L., Golberg R. (1982). Structure and flanking regions of soybean seed protein genes. *Cell.* 29: 651-
- Foster R., Izawa T., Chua N.-H. (1994) Plant bZIP proteins gather at ACGT elements. *FSEB J.* 8: 192-200.

- Fujiwara T., Beachy R.N. (1994) Tissue-specific and temporal regulation of a β -conglycinin gene: roles of the RY repeat and other *cis*-acting elements. *Plant Mol. Biol.* 24: 261-272.
- Galau G.A., Hughes D.W., Dure L. III. (1986) Absciscic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. *Plant Mol. Biol.* 7: 155-170.
- Galau G.A., Jakobsen K.S., Hughes D.W. (1991) The controls of late dicot embryogenesis and early germination. *Physiol. Plant.* 81: 280-288.
- Gatehouse J.A., Shirsat A.H. (1993) Control of expression of seed storage protein genes In: Ed. Verma DPS *Control of Plant Gene Expression*, 357-392. CRC Press, Inc., Boca Raton, FL.
- Gaubier P., Raynal M., Hull G., Huestis G.M., Grellet F., Arenas C., Pages M., Delseny M. (1993) Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Mol Gen Genet* 238: 409-418.
- Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F., Goodman H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4: 1251-1261.
- Goldberg R.B., Barker S.J., Perez-Grau L. (1989) Regulation of gene expression during plant embryogenesis. *Cell* 56: 149-160.
- Goldberg R.B., Paiva G.d., Yadegari R. (1994) Plant embryogenesis: zygote to seed. *Science* 266: 605-614.
- Guiltinan M.J., Marcotte J.W.R., Quatrano R.S. (1990) A plant leucine zipper protein that recognizes an absciscic acid response element. *Science* 250: 267-271.
- Hammond-Kosack M.C.U., Holdsworth M.J., Bevan M.W. (1993) In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J.* 12: 545-554.
- Hartings H., Maddaloni M., Lazzaroni N., Fonzo N.D., Motto M., Salamini F., Thompson R. (1989) The O2 gene which regulated zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J.* 8: 2795-2801.
- Hattori T., Vasil V., Rosenkrans L., Hannah L.C., McCarty D.R., Vasil I.K. (1992) The *Viviparous-1* gene and absciscic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev.* 6: 609-618.
- Hughes D.W., Galau G.A. (1989) Temporally modular gene expression during cotyledon development. *Genes Dev.* 3: 358-369.

- Hughes D.W., Galau G.A. (1991) Developmental and environmental induction of Lea and LeaA mRNAs and the postabscission program during embryo culture. *Plant Cell* 3: 605-618.
- Izawa T., Foster R., Nakajima M., Shimamoto K., Chua N.-H. (1994) The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. *Plant Cell* 6: 1277-1287.
- Jakobsen k.S., Hughes D.W., Galau G.A. (1994) Simultaneous induction of postabscission and germination mRNAs in cultured dicotyledonous embryos. *Planta* 192: 384-394.
- Jofuku K.D., Schipper R.D., Goldberg R.B. (1989) A frameshift mutation prevents kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *Plant Cell* 1: 427-435.
- Jordano J., Almoguerra C., Thomas T.L. (1989) A sunflower helianthinin gene upstream sequence ensemble an enhancer and sites of nuclear protein interaction. *Plant Cell* 1: 855-866.
- Karssen C.M., Swan D.L.C.B.-v.d., Breekland A.E., Koornneef M. (1983) Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of *Arabidopsis thaliana* (L.) Heynh. *Planta* 157: 158-165.
- Kawagoe Y., Campell B.R., Murai N. (1994) Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein β -phaseolin gene. *Plant J.* 5: 885-890.
- Kawagoe Y., Murai M. (1992) Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein b-phaseolin gene conferring spatial and temporal control. *Plant J.* 2(6): 927-936.
- Keith K., Kraml M., Dengler N.G., McCourt P. (1994) *fusca3*: a heterochronic mutation affecting late embryo development in *Arabidopsis*. *Plant Cell* 6: 589-600.
- Koornneef M., Hanhart C.J., Hilhorst H.W.M., Karssen C.M. (1989) *In vivo* inhibition of seed development and reserve protein accumulation in recombinants of abscisic acid biosynthesis and responsiveness mutants in *Arabidopsis thaliana*. *Plant Physiol.* 90: 463-469.
- Koornneef M., Jorna M.L., Swan D.L.C.B.-v.d., Karssen C.M. (1982) The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* 61: 385-393.

- Koornneef M., Reuling G., Karssen C.M. (1984) The isolation and characterization of abscisic acid insensitive mutants of *Arabidopsis thaliana*. *Physiol. Plant.* 61: 377- 383.
- Kriz A.R., Wallace M.S., Paiva R. (1990) Globulin gene expression in embryos of maize viviparous mutants. *Plant Physiol.* 92: 538-542.
- Lam E., Chua N.-H. (1991) Tetramer of a 21-base pair synthetic element confers seed expression and transcriptional enhancement in response to water stress and abscisic acid. *J. Biol. Chem.* 266: 17131-17135.
- Lelievre J.-M., Oliveira L.O., Nielsen N.C. (1992) 5'-CATGCAT-3' elements modulate the expression of glycinin genes. *Plant Physiol.* 98: 387-391.
- Lessard P.A., Allen R.D., Bernier F., Crispino J.D., Fujiwara T., Beachy R.N. (1991) Multiple nuclear factors interact with upstream sequences of differentially regulated β -conglycinin genes. *Plant Mol. Biol.* 16: 397-413.
- Lessard P.A., Allen R.D., Fujiwara T., Beachy R.N. (1993) Upstream regulatory sequences from two β -conglycinin genes. *Plant Mol. Biol.* 22: 873-885.
- Litts J.C., Colwell G.W., Chakerian R.L., Quatrano R.S. (1987) The nucleotide sequence of a cDNA clone encoding the wheat E_m protein. *Nucl. Acids Res.* 15: 3607-3618.
- Lohmer S., Maddaloni M., Motto M., Di Fonzo N., Hartings H., Salamini F., Thompson R.D. (1981) The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of *b-32* gene. *EMBO J.* 10:617-624.
- Lopes M.A., Larkins B.A. (1993) Endosperm origin, development, and function. *Plant Cell* 5: 1338-1399.
- Lüscher B., Eisenman R.N. (1990) New light on Myc and Myb. Part I. Myc. *Genes and Development* 4: 2025-2035.
- Marcotte Jr.W.R., Russell S.H., Quatrano R.S. (1989) Absciscic acid-responsive sequences from the *Em* gene of wheat. *Plant Cell* 1: 969-976.
- Marks G.E., Davies D.R. (1969) The cytology of cotyledon cells and the induction of giant polytene chromosomes in *Pisum sativum*. *Protoplasma* 101: 73-81.
- McCarty D.R., Carson C.B., Stinard P.S., Robertson D.S. (1989) Molecular analysis of viviparous-1: An abscisic acid-insensitive mutant of maize. *Plant Cell* 1: 523-532.

- McCarty D.R. and Carson C.H. (1991) The molecular genetics of seed maturation in maize. *Physiol. Plant.* 81: 267-272.
- McCarty D.R., Hattori T., Carson C.B., Vasil V., Lazar M., Vasil I.K. (1991) The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66: 895-905.
- McCubbin W.D., Kay C.M., Lane B.G. (1985) Hydrodynamic and optical properties of the wheat germ E_m protein. *Can. J. Biochem. Cell Biol.* 63: 803-811.
- Mundy J., Chua N.H. (1988) Absciscic acid and water-stress induce the expression of a novel rice gene. *EMBO J.* 7: 2279-2286.
- Mundy J., Yamaguchi-Shinozaki K., Chua N.-C. (1990) Nuclear proteins bind conserved elements in the absciscic acid-responsive promoter of a rice *rab* gene. *Proc. Natl. Acad. Sci. USA* . 87: 1406-1410.
- Naito S., S. Hirai M.Y., Chino M., Komeda Y. (1994) Expression of a soybean (*Glycine max* [L.] Merr. seed storage protein gene in transgenic *Arabidopsis thaliana* and its response to nutritional stress and to absciscic acid mutations. *Plant. Physiol.* 104: 497-503.
- Nambara E., Keith K., McCourt P., Naito S. (1995) A regulatory role for *ABI3* gene in the establishment of embryo maturation in *Arabidopsis thaliana*. *Development* 121: 629-636.
- Paiva R., Kriz A.L. (1994) Effect of absciscic acid on embryo-specific gene expression during normal and precocious germination in normal and *viviparous* maize (*Zea mays*) embryos. *Planta* 192: 332-339.
- Parcy F., Valon C., Raynal M., Gaubier-Comella P., Delseny M., Giraudat J. (1994) Regulation of gene expression programs during *Arabidopsis* seed development: Role of the *ABI3* locus and of endogenous absciscic acid. *Plant Cell* 6: 1567-1582.
- Perez-Grau L., Goldberg R.B. (1989) Soybean seed protein genes are regulated spatially during embryogenesis. *Plant Cell* 1: 1095-1109.
- Pla M., Gomez J., Goday A., Pages M. (1991) Regulation of the absciscic acid-responsive gene *rab28* in maize *viviparous* mutants. *Mol Gen Genet* 230: 394-400.
- Pla M., Vilardell J., Gultinan M.J., Marcotte Jr.W.R., Niogret M.F., Quatrano R.S., Pages M. (1993) The *cis*-regulatory element CCACGTGG is involved in ABA and water-stress reponses of the maize gene *rab28*. *Plant Mol. Biol.* 21: 259-266.

- Pysh L.D., Aukerman M.J., Schmidt R.J. (1993) OHP1: A maize basic domain/leucine zipper protein that interacts with Opaque-2. *Plant Cell* 5: 227-236.
- Quatrano R.S. (1986) Regulation of gene expression by abscisic acid during angiosperm embryo development. *Oxford Surveys Plant. Mol. Cell Biol.* 3: 467-477.
- Quatrano R.S., Marcotte Jr.W. R., Guiltinan M. (1993) Regulation of gene expression by abscisic acid In: Ed. Verma DPS *Control of Plant Gene Expression*, 69-90. CRC Press, Boca Raton, FL.
- Quayle T., Feix G. (1992) Functional analysis of the -300 region of the maize zein genes. *Mol. Gen. Genet.* 231: 369-374.
- Robertson D.S. (1955) The genetics of vivipary in maize. *Genetics* 40: 745-760.
- Robichaud C.S., Wong J., Sussex I.M. (1980) Control of *in vitro* growth of viviparous embryo mutants of maize by abscisic acid. *Developmental Genetics* 1: 325-330.
- Salinas J., Oeda K., Chua N.-H. (1992) Two G-box-related sequences confer different expression patterns in transgenic tobacco. *Plant Cell* 4: 1485-1493.
- Schmidt R.J., Burr F.A., Aukerman M.J., Burr B. (1990) Maize regulatory gene *Opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc. Natl. Acad. Sci. USA.* 87:46-50.
- Schmidt R.J. (1993) Opaque-2 and zein gene expression In: Ed. Verma DPS *Control of Plant Gene Expression*, 337-355. CRC Press, Inc., Boca Raton, FL.
- Schmidt R.J., Ketudat M., Aukerman M.J., Hoschek G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-KD zein genes. *Plant Cell* 4: 689-700.
- Shen Q., Ho T.-H.D. (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. *Plant Cell* 7: 295-307.
- Shen Q., Ukness S.J., Ho T.-H.D. (1993) Hormone response complex of a novel abscisic acid and cycloheximide inducible barley gene. *J. Biol. Chem.* 268: 23652-23660.
- Skriver K., Mundy J. (1990) Gene expression in response to abscisic acid and osmotic stress. *Plant Cell* 2: 503-512.

- Thomas T.L. (1993) Gene expression during plant embryogenesis and germination. *Plant Cell* 5: 1401-1410.
- Ueda T., Wawerczak W., Ward K., Sher N., Ketudat M., Schmidt R., Messing J. (1992) Mutations of the 22- and 27-KD zein promoters affect transactivation by the Opaque-2 protein. *Plant Cell* 4: 701-709.
- Ulrich T.U., Wurtele E.S., Nikolau B.J. (1990) Sequence of EMB-1, an mRNA accumulating specifically in embryos of carrot. *Nucl. Acids Res.* 18: 2826.
- van der Geest A.H.M., Hall G.E.J., Spiker S., Hall T.C. (1994) The β -phaseolin gene is flanked by matrix attachment regions. *Plant J.* 6: 413-423.
- Vetten N.C.d., Lu G., Ferl R.J. (1992) A maize protein associated with the G-box binding complex has homology to brain regulatory proteins. *Plant Cell* 4: 1295-1307.
- Walling L., Drews G.N., Goldberg R.B. (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc. Natl. Acad. Sci. USA* 83: 2123-2127.
- West M.A.L., Harada J.J. (1993) Embryogenesis in higher plants: an overview. *Plant Cell* 5: 1361-1369.
- West M.A.L., Yee K.M., Danao J., Zimmerman J.L., Fischer R.L., Goldberg R.B. (1994) *LEAFY COTYLEDON1* is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* 6: 1731-1745.
- Williamson J. D., Scandalios J. (1992) Differential response of maize catalases to abscisic acid: Vp1 transcriptional activator is not required for abscisic acid-regulated *Cat1* expression. *Proc. Natl. Acad. Sci. USA.* 89: 8842-8846.
- Wilcox J.R. (1987) Soybeans: Improvement, Production, and Uses. American Society of Agronomy, Madison, WI.,
- Yadegari R., de Paiva G.R., Laux T., Koltunow A.M., Apuya N., Zimmerman J.L., Fischer R.L., Harada J.J., Goldberg R.B. (1994) Cell differentiation and morphogenesis are uncoupled in *Arabidopsis* raspberry embryos. *Plant Cell* 6: 1713-1719.
- Yunes J.A., Neto G.C., Silva M.J.d., Leite A., Ottoboni L.M.M., Arruda P. (1994) The transcriptional activator opaque2 recognizes two different target sequences in the 22-kD-like α -prolamin genes. *Plant Cell* 6: 237-249.

**CHAPTER 1: CLONING, MAPPING, AND ANALYSES OF EXPRESSION
OF THE *Em*-LIKE GENE FAMILY IN SOYBEAN (*Glycine max* (L).
Merr.)***

A paper submitted to Plant Molecular Biology

Éberson S. Calvo, Eve S. Wurtele, and Randy C. Shoemaker

Abstract

In this study we cloned the entire *Em*-like Group 1 late embryogenesis abundant (Lea) gene family in soybean. In support of the hypothesis of the tetraploid origin of the soybean genome, the five Group 1 Lea genes (*S/e*1-5) could be divided into two classes based on sequence identity. *S/e*1-4 were placed on the soybean genetic map and were shown to map to 4 different linkage groups. *S/e*1, *S/e*2, *S/e*3, and *S/e*5 encode polypeptides differing primarily by the presence of a repeated 20-amino acid motif. *S/e*1 and *S/e*5 were shown to be expressed in developing embryos weeks earlier than *S/e*2 and *S/e*3. *S/e*4 was shown to be a pseudogene. Maximal levels of mRNA for all functional *S/e* genes accumulated in maturation phase seeds, before significant desiccation had occurred, and declined rapidly upon seed imbibition. Desiccation did not induce *S/e* expression in seeds. ABA was not necessary to induce *S/e* expression in cultured embryos and only moderately increased *S/e* mRNA accumulation. *S/e* expression was confined to embryo tissues and *S/e* mRNA accumulated at similar levels in both the embryo

* The DNA sequences presented in this paper are available in the EMBL/GenBank databases under the accessions numbers:

axis and in the cotyledons. Neither ABA nor desiccation induced expression of the *S/e* genes in vegetative tissues.

Introduction

Due to its worldwide agronomic importance the physiology and biochemistry of soybean seed development have been extensively studied. Recently, a considerable amount of information has accumulated on the molecular biology of the events involved in the developmental regulation of expression of several seed protein genes, particularly those encoding storage proteins [19,20]. Among the concepts derived from these studies is that gene expression during seed development is temporally and spatially regulated at both the transcriptional and postranscriptional levels [19,29,45,59].

The hormone ABA has been implicated as a positive regulator of gene expression in seeds of different plant species [15,16,22,26,44,47,53]. These observations were made for genes belonging to the different temporal programs described for embryos [25,27], including several storage protein genes [54] expressed during the maturation program, and a variety of genes expressed during late embryogenesis, the so called late embryogenesis abundant (Lea) genes [15]. ABA has been shown to regulate the expression of the β -conglycinin gene of soybean in *in vitro* cultured embryos [5] and in transgenic *Arabidopsis* plants [41]. To date no reports have been made on ABA-mediated regulation of any of the Lea gene(s) in soybean.

Despite this evidence of ABA-mediated gene expression, the role of endogenous ABA as the developmental regulator of Lea gene expression in dicot species has been questioned [16,26]. Recent studies with *Arabidopsis* mutants

have shown that although ABA may act as an additional modulator of expression for some Lea genes, not all Lea genes are affected by the endogenous ABA concentration [44,56]. Consequently, developmental factor(s) other than ABA must exist in the *Arabidopsis* embryos which are able to elicit the same temporal response.

Much information on the regulation of Lea gene expression has been derived from experiments using monocot species [37,38,39]. However, this information can not necessarily be extrapolated to dicots [16]. Unlike monocots, where storage protein gene expression is frequently concentrated in the endosperm, the embryo cotyledons and to some extent the embryo axis are the primary tissue for storage protein gene expression in seeds of the majority of dicot species [36]. This is particularly true for legumes such as soybean where the endosperm is very short lived [7] and where storage proteins account for 36 % or more of the seed dry weight [62]. As a consequence, in soybean, expression of both the seed-specific maturation and late embryogenesis programs may take place exclusively in the embryo. Therefore, it is conceivable that embryos of legume species may utilize a somewhat different mechanism from monocots to achieve the temporally distinct expression of the maturation and late embryogenesis programs.

The Lea genes can be divided into groups based on the structure of the encoded protein [8]. The *Em* gene from wheat [33,34] was the first Group 1 [8] Lea gene isolated from plants. Isolation of homologous genes or cDNAs have been reported from eight other species [2,3,10,18,35,49,55,60,63]. Nevertheless, none of these reports include genes or cDNAs from legume species.

Our primary objective was to clone, sequence and analyze the expression of the complete *Em*-like gene family from soybean as the framework to study gene

family evolution and regulation of Lea gene expression during soybean seed development. This report identifies and characterizes all five members of the Group 1 Lea gene family from soybean, and investigates the role of ABA relevant to Group 1 Lea gene expression.

Materials and Methods

Plant Material

Glycine max (L.) Merr. cv Williams 82 was used as the source of immature seeds obtained from plants grown under standard greenhouse conditions. Flowers were tagged at anthesis and immature seeds were collected at various days after flowering (DAF). Fifty seeds of approximately the same size were pooled for each time point, their total fresh weight was recorded, and the seeds were immediately frozen in liquid nitrogen, and stored at -80 °C. Leaf tissue was obtained from two-week-old plants that were grown in a sandbench in a greenhouse. Seeds were germinated in the dark at 27 °C between two rolled sheets of germination paper, which had been saturated by emersion in sterile water. At timed intervals after imbibition, seedling tissues were frozen in liquid nitrogen and stored at -80 °C.

ABA and Water Deficit Treatments

For the ABA treatment of embryos, zygotic embryos at 30 DAF were excised from plants and cultured on MS [40] medium with 0 or 100 μ M ABA for 24 h. For the water deficit treatments, seeds containing embryos at 30 DAF were left in excised pods and allowed to dry for 48 h [50]. For both the ABA and the water

deficit treatments, treated embryos (1 to 2 g fresh weight) were harvested, frozen in liquid nitrogen, and stored at -80°C.

For the ABA treatment of seedlings, four-day-old seedlings were transferred to a petri dish containing MS [40] medium, with or without 100 μ M ABA, for two days. For the water deficit treatments, germination papers containing four-day-old seedlings were removed from the water-containing tray, unfolded, laid over dry trays, and kept in a cabinet for 26 h to ensure a slow dehydration. Tissues of the seedlings were visibly dehydrated as judged by the turgor of the hypocotyl. Seedlings in which the root tip was completely dry were discarded. Roots, hypocotyls, and cotyledons of 10 to 15 seedlings were harvested and immediately frozen in liquid nitrogen and stored at -80 °C.

For the ABA treatment of plants, five-week-old greenhouse-grown plants were sprayed with water (control) or with a solution containing 100 μ M ABA. Plants were sprayed twice a day (at 8 A.M. and 6 P.M.) for two days. Tween 20 (0.05 % [v/v]) was included as a surfactant for both treatments. Forty-eight h after the initial spraying, leaflets from 12 plants were harvested, pooled, frozen in liquid nitrogen, and stored at -80 °C. For the water deficit treatments, shoots of five-week-old plants were harvested, weighed, and left on the laboratory bench for about 6.5 h, until shoots had lost 20 % of their initial weight [46]. Leaves from 12 plants were harvested, pooled, frozen in liquid nitrogen, and stored at -80 °C.

Isolation of cDNA and Genomic Clones

A λ gt11 soybean (cv. Enrei) cDNA library, constructed from mid-maturation developing seeds, was kindly provided by Drs. Daisuke Shibata and Robert Whittier (Mitsui Plant Biotechnology Research Institute, Tsukuba, Japan).

Bacteriophage manipulation and screening techniques were as described by Sambrook et al. (1989). Nitrocellulose (Schleicher & Schuell) filters were pre-hybridized and hybridized in 6 X SSC (1 X SSC is 0.18 M NaCl and 1.9 mM sodium citrate, pH 7.0), 1 % (w/v) SDS, 25 mM NaHPO₄, pH 6.5, 3 X Denhardt's, and 0.1 mg/mL herring sperm DNA, at 65 °C. Final washes were done in 0.5 X SSC, 0.1 % SDS, at 56°C, for 1 h. The probe used for screening the library was the carrot EMB-1 cDNA [55] labeled with [α -³²P]-dCTP [11]. The cDNAs hybridizing to the EMB-1 probe were subcloned into the EcoRI site of pBluescript KS⁺ (Stratagene).

Genomic clones for *S/e1* and *S/e4* were isolated from a λ EMBL3 genomic library made from cv. Williams (Clontech). Genomic clones for *S/e2* and *S/e3* were isolated from a λ EMBL3 library made from cv. Resnik (Clontech). The *S/e5* clone was isolated as a 2.5 Kb EcoRI fragment from a λ gt10 genomic library made after gel fractionating (2.3 - 3.5 Kb) EcoRI digested DNA (cv. Williams) in low melting agarose (FMC). DNA was recovered from the gel by using a QIAGEN tip-5 column essentially as described by the manufacturers. Library construction followed procedures described in Sambrook et al, (1989). The three genomic libraries were screened with the soybean cDNA. Screening conditions were as described for cDNA library except that final wash was done at 2 X SSC, 0.1 % SDS, at 50 °C.

DNA Sequencing and DNA Comparisons

Automated sequencing was performed on both single and double strand DNA templates in an ABI sequencing apparatus at the Iowa State University Nucleic Acid Facility. Manual sequencing was performed according to the dideoxynucleotide chain termination method [51] using a Sequenase 2.0 kit from

USB. All sequence analyses were performed with the Mac DNAsis sequence analysis software (Hitachi) using default settings in the Higgins-Sharp mode. The 3'-UTR regions used for sequence comparisons comprise a 520 bp fragment immediately downstream of the stop codon of each clone. For sequence comparison of the coding regions we omitted the DNA sequence corresponding to the second 20-amino acid hydrophilic motif from clones *S/e1* and *S/e5*. Intron comparisons were made with the entire intron sequences.

RFLP Mapping

The map locations of four *S/e* clones were determined by restriction fragment length polymorphism analysis of the F2 segregating population described by Keim et al. (1990) using the MapMaker software [31]. Identity of each locus was assigned based on fragment sizes predicted from the restriction map obtained for individual λ EMBL3 clones.

Isolation of DNA and RNA, and Hybridization Analyses

Soybean leaf DNA was isolated following procedures described by Keim et al. (1990). Total RNA was isolated from different tissues and at different developmental stages by using the mini-prep procedure described by Wadsworth et al. (1988), with an extraction buffer containing guanidine thiocyanate as a chaotropic agent. A chloroform extraction was included immediately after the phenol:chloroform step in order to better remove lipids from seed tissues.

DNA was digested with different restriction enzymes, fractionated by electrophoresis in 0.8 % (w/v) agarose gels, and transferred to nylon membranes

(Biotrace, Gelman Sciences) according to standard procedures [52]. Pre-hybridizations and hybridizations were performed as described. Final washes were done in 0.5 X SSC, 0.1% (v/v) SDS, at 65 °C for 40 min.

Hybridization analyses of RNA blots were performed according to Sambrook et al. (1989). RNA was fractionated in 1% (w/v) agarose gels containing 6% (v/v) formaldehyde. All RNA samples were denatured in the presence of 1 µg ethidium bromide. Nylon membranes were pre-hybridized in a solution containing 50% (v/v) formamide, 5 X SSC, 2 X Denhardt's, 40 mM NaHPO₄, pH 6.5, 10 mM EDTA, and 0.2 mg/mL herring sperm DNA. After 12 h of pre-hybridization, a [³²P]-labeled probe was added, and hybridization continued for 15 to 18 h. Final washes were carried out under conditions given in the figures. The soybean actin cDNA probe, pSAC-7 (kindly made available by Dr. Richard Meagher, Univ. of Georgia, Athens), was used as an internal control in the RNA blots. All RNA blots were performed twice with replicated samples from two independent extractions. DNA probes were labeled by random-priming [11]. *S/e1* and *S/e5* 3'-UTR gene-specific probes comprise the fragments spanning from positions 379 to 687 bp and 1201 to 1740 bp, on the *S/e1* cDNA and *S/e5* genomic clones, respectively. Oligonucleotide gene-specific probes were hybridized in the absence of Denhardt's solution at T_m-10 °C. The oligonucleotides used were: 5'-GGTCTTCGTTTCTGATTCTGGTT-3' and 5'-GACCTAGCTGCTACCTATACCAT-3' for *S/e2* and *S/e3*, respectively. Oligonucleotides were end-labeled using polynucleotide kinase as described in Sambrook et al. (1989).

Reverse Transcription and PCR Reactions

One microgram of total RNA, prepared as described above from 87 DAF seeds, was reverse transcribed using a first-strand cDNA kit (Pharmacia), essentially as described by the manufacturers. One hundredth of the reaction (or 50 ng of soybean genomic DNA in the control reactions) was used as template in a 100 μ l PCR reaction (200 μ M dNTP, 0.2 μ M of each primer, 1 X PCR buffer (Promega), 2 mM $MgCl_2$, 1 U Taq polymerase (Promega)) using the combination of primers described in the text. Sequences of the primers used were as follow : P0 (5'-CTTGAGGCTCAAGAACAT CTTGCTG); P1/4 (5'- ACCAGACTTGTCCATGGTG CTGAG-3'). A total of 40 PCR cycles were performed using the following conditions: 45" at 93 °C, denaturing step; 2' at 70 °C, annealing step; and 2' at 72 °C, extension step.

Results

Screening of the cDNA Library

Over 500,000 recombinant bacteriophage from the soybean cDNA library were screened by hybridization with the radioactively labeled EMB-1 cDNA of carrot [55]. This screen resulted in the selection of 8 soybean cDNA clones. Restriction mapping and partial DNA sequencing of the soybean clones suggested that all of these clones contained identical cDNAs. For this reason, only one clone, Sle1 (soybean late embryogenic), was examined in detail.

The cDNA clone Sle1 (GenBank accession number) is 676 bp in length, excluding the poly A tail, and contains a translational open reading frame (ORF)

that codes for a polypeptide of 112 amino acids. This is the longest ORF and its first ATG is preceded by an in frame stop codon. The deduced Sle1 polypeptide shows all the characteristics of a Group 1 Lea protein [8]. Sequence comparison with all previously described Group 1 Lea proteins revealed that Sle1 has on average 75 % amino acid identity and 85 % similarity to its homologs. The Sle1 polypeptide also contains a duplicated, highly hydrophilic, 20-amino-acid motif that has been previously found in the barley, cotton, and *Arabidopsis* homologs. [10,18].

Genetic Mapping of the *S/e* Gene Family

DNAs isolated from *G. max* (A81-356022) and *G. soja* (PI 468.916) were subjected to restriction endonuclease digestion and analyzed by hybridization to the Sle1 cDNA probe. The Sle1 cDNA hybridized to several (3 to 6) genomic fragments for all of the restriction enzymes tested (Figure 1). This observation is in agreement with other studies indicating that most Group 1 Lea genes appear to be members of multigene families [10,14,18,33].

In order to identify all of the members of the *S/e* gene family in soybean, we screened two independent λ EMBL3 genomic libraries under low stringency conditions (final wash in 2 X SSC, 0.1 % SDS, 50°C). Extensive restriction mapping of several clones enabled us to identify four different non-overlapping genomic clones containing *S/e*-hybridizing fragments. A detailed restriction map of these four clones is presented in Figure 2. The clones were designated *S/e*1-4, with *S/e*1 being the clone that most resembles the Sle1 cDNA restriction map (this was later confirmed by sequence analyses). These four clones could explain all but one of the 6 Hind III and 5 EcoRI fragments observed in Figure 1. Therefore, we

also screened an EcoRI size-fractionated library constructed in λ gt10 vector. This screening allowed us to isolate a fifth, *S/e* 5, clone as a 2.5 Kb EcoRI fragment (Figure 2).

Figure 1 shows that the *S/e*-hybridizing fragments are polymorphic between the two soybean species tested. RFLP mapping of the two *Dra*I and the EcoRI polymorphisms displayed in Figure 1 revealed three independent *S/e* loci mapping to linkage groups D1, G, and N of the USDA-ARS soybean RFLP linkage map (Figure 3). Based on the restriction map of the clones and the fortuitous absence of polymorphisms among the cv. Resnik, cv Williams, and A81-356022 G. max genotypes (data not shown), we were able to assign a specific clone to each one of the mapped polymorphisms. The fourth λ EMBL3 clone, *S/e*4, was also mapped to the RFLP linkage map by mapping a low-copy *Taq*I fragment contiguous to the *S/e* hybridizing region (Figure 2). The mapped positions of the clones are presented in Figure 3. All four *S/e* genes map to independent linkage groups. We were unable to unambiguously map the fifth clone, *S/e*5, but preliminary data indicates that this locus may map to a fifth linkage group (unpublished results). In conclusion, the *S/e* genes in soybean comprise a small gene family with five members that map to four, and possibly five, different chromosomes in the soybean genome.

Sequence Analysis of Genomic Clones

We further characterized the *S/e* gene family by sequencing each of the five *S/e* genomic clones. Figure 4 shows the sequence of the five *S/e* genes and their deduced polypeptides. Since we have characterized a cDNA clone only for the *S/e*1 locus, the deduced polypeptides and intron positions for the remaining loci (*S/e*2-5) are only predicted. Nevertheless, protein sequence and intron locations

have been found to be very conserved among the *S/e* homologs in other species [2,3,10,18,35,49,55,60,63]. Results from DNA sequence comparison among the five genes are presented in Table I. Sequence identity throughout the coding regions was very high among all clones. However, the clones can be divided into two classes based on the sequence identity of the coding regions. *S/e2* and *S/e3* are more related to each other (87.2 % identity) than to the remaining clones (average of 72.9 % identity). *S/e1*, *S/e4*, and *S/e5* comprise the second class of clones, showing higher homology to each other (average of 88.4 % identity) than to either clone *S/e2* or *S/e3*. Comparison of the deduced polypeptide sequences of each *S/e* locus (partial sequence for *S/e5* polypeptide) shows that they differ primarily in that *S/e1* and *S/e5* have a repeated 20 amino acid hydrophilic motif. The *S/e4* locus encodes a polypeptide containing a premature stop codon and therefore is likely to be a pseudogene.

All five clones appear to have a single intron (Figure 4), as has been found for other *S/e* homologs [17,18,23,34,35,63]. *S/e5* has an unusually large intron; over 1700 bp (Figure 4; data not shown). We were unable to determine the complete sequence of the *S/e5* intron because the 2.5 Kb fragment comprising the *S/e5* clone terminates at its 5' end within the intron. Predicted Intron lengths for clones *S/e1*, 2, 3 and 4 were 250, 509, 811 and 249 bp, respectively (Figure 4). Intron lengths from previously sequenced *S/e* homologs in six different plant species have ranged from 71 bp in wheat [34] to 185 bp in *Arabidopsis* [18]. Sequence identity of *S/e1*, *S/e4*, and *S/e5* putative introns averaged 52.7 % (Table I). *S/e4* and *S/e5* introns showed the highest identity (65 %) differing almost exclusively by the presence of three insertions in the longer *S/e5* intron. *S/e2* and *S/e3* introns showed 44 % identity. Sequence identity across introns from the two

classes was found to be very low (31 %) and therefore also corroborate the existence of two different classes.

Sequences downstream of the translation stop site of each gene (Table I) have a high overall sequence identity within each class, averaging 57.7 % for *Sle2* and *Sle3*, and 54.1 % for the three remaining genes. When the comparison is made across the two classes, this value drops to 24.8 %. As observed for the introns, *Sle4* and *Sle5* 3'-UTR are more identical to each other than to *Sle1* 3'-UTR.

Sle mRNA Accumulation in Embryos

Expression of Group 1 Lea genes in other plant species has been shown to occur predominantly in embryonic tissues [18,60,63]. To investigate the tissue specificity and the pattern of expression of the *Sle* gene family, we examined *Sle* mRNA accumulation by northern analyses of total RNA extracted from vegetative tissues and developing seeds of soybean. We initially approached this question by using the *Sle1* cDNA as a probe under low stringency conditions (final wash in 2 X SSC, 0.1 % SDS at 50 °C). This probe cross-hybridizes with all other *Sle* mRNA under these conditions (data not shown). No hybridization was detected in RNA from leaf, stem, or root tissues (data not shown). A single band corresponding to mRNA of approximately 850 bp was detected in seed tissues (Figure 5). To better localize the accumulation of *Sle* mRNA in the seed, we dissected mature dry seeds and analyzed total RNA isolated from the embryonic axis and from the cotyledons by RNA hybridization analysis. Figure 5 shows that approximately equal amounts of *Sle* mRNA accumulate in the two embryonic tissues.

The steady-state levels of *Sle* mRNA decreased rapidly after imbibition (Figure 6), as do other Group 1 *Lea* genes [49,60]. By 32 h post-imbibition, *Sle* mRNA could not be detected in the seedling. In contrast, actin mRNA accumulated during germination. The pattern of actin mRNA accumulation was the opposite of that observed for *Sle* mRNA, with increasing actin mRNA accumulation from 0 to 80 h post-imbibition.

Next, we investigated in more detail the pattern of *S/e* gene expression throughout soybean seed development (Figure 7). Hybridization with the *Sle1* cDNA probe shows that *Sle* mRNA could be detected at relatively low levels in 25 DAF seeds which have already entered the early cotyledonary stage of seed development. The *Sle* mRNA steady-state level remains constant throughout seed development until a noticeable increase occurs around 80 DAF, at a time when seeds have reached maximal size and are starting to turn yellow in the region surrounding the embryo-axis. At 87 DAF maximum levels of *Sle* mRNA are detected. At 105 DAF, seeds were fully matured and dried and levels of *Sle* mRNA were still high. *Sle* mRNA reaches its maximum levels before any appreciable loss of fresh weight has occurred in the seed. In contrast, expression of the β -conglycinin storage protein peaks much earlier than *S/e* and is steadily decreasing by the time *S/e* expression starts to increase.

Our next question was to determine whether or not all the individual *S/e* genes had similar patterns of expression. For this purpose we used gene-specific probes (see Methods). The results on Figure 7 show that although all four genes are being expressed, their temporal pattern of expression differs considerably, with *Sle1* and *Sle5* being expressed much earlier in seed development than *Sle2* and *Sle3*.

Sequence analysis had indicated that *Sle4* is a pseudogene. To confirm that indeed *Sle4* mRNA was not present we performed RT-PCR [13] analysis on developing seeds using a set of primers which would specifically amplify mRNA derived from *Sle1* (as an internal control of the RT-PCR reactions) and *Sle4* loci (Figure 8A). Since these two clones differ by the presence of a 20-amino-acid repeat, we predicted that their RT-PCR products would differ by 60 bp (199 and 139 bp, respectively; Figure 8A). Results of this experiment show that only one band can be seen after PCR amplification (Figure 8B, lane 4). The band size corresponds to the expected amplification product from a *Sle1* derived mRNA (199 bp). That the set of primers and reaction conditions chosen are adequate to amplify the *Sle4* gene is shown by the control reactions performed with genomic DNA as template (Figure 8B, lanes 1 and 3), where the bands corresponding to amplification of the two genes are present. Therefore, the absence of the expected *Sle4* mRNA band in the RT-PCR reaction is a strong evidence that *Sle4* mRNA does not accumulate at appreciable levels in the cell.

Sle mRNA Accumulation in Response to ABA and Water Deficit

The expression of Group 1 *Lea* genes in cultured zygotic embryos of cotton [26], wheat [61], maize, [60], and barley [10] is strongly increased by the addition of exogenous ABA. To examine whether *Sle* mRNA accumulation in zygotic embryos is induced by ABA and/or water deficit, and, if so, to determine whether such an induction is similar at different developmental stages, zygotic embryos, four-day-old seedlings, and five-week-old plants of soybean were subjected to ABA and water deficit and analyzed for *Sle* mRNA accumulation under low stringency conditions.

Zygotic embryos (30 DAF) were removed from the pod and cultured for 24 h in medium with or without added ABA. The transfer of embryos to culture medium for 24 h significantly increased *Sle* mRNA accumulation even in the absence of exogenous ABA (Figure 9A, compare lanes 1 and 2). The presence of ABA in the culture medium only slightly increased *Sle* mRNA levels (Figure 9A, compare lanes 2 and 3).

Seed desiccation, rather than ABA, has been suggested as the requirement for completion of seed maturation [12], and thus *Lea* gene expression may be associated with this physiological change. However, desiccation of the embryos in the pod, as described by Rosenberg and Rinne (1988), resulted in a reduction of *Sle* mRNA accumulation in 30 DAF seeds (Figure 9A, lane 4).

In contrast to embryos, vegetative tissues did not respond either to ABA or water deficit by accumulation of *Sle* mRNA. Neither ABA nor water deficit induced *Sle* mRNA accumulation to detectable levels in leaves of five-week-old plants (Figure 9B). Furthermore, cotyledons, hypocotyls, and roots of four-day-old seedlings did not accumulate *Sle* mRNA to detectable levels in the presence of ABA, or when subjected to water deficit (data not shown).

Discussion

The gene family presented here represents the largest *Em*-like gene family described in plants and the first among legume species. Based on sequence identity, we grouped the five members of the *S/e* gene family into two classes. It has been suggested that soybean is an ancient allotetraploid, and putative homeologous genes have been molecularly characterized for some gene families in soybean [21,32,42]. Therefore, one possibility is that the *S/e* genes belonging to

the same class have originated from the two different genomes that came together to form what we currently know as soybean. Thus, each genome would have a copy of the differentially regulated genes. However, since the two ancestral genomes are not known, we can not unambiguously test this hypothesis.

Sequence and RT-PCR analyses showed that *S/e4* is a pseudogene and *S/e4* mRNA does not accumulate in the cell. We do not know whether *S/e4* is being transcribed. Nevertheless, frameshift mutations that lead to introduction of premature stop codons have been held responsible for the absence of mRNA accumulation in soybean [28] and *Phaseolous vulgaris* [57]. Therefore, it is possible that the nonsense point mutation in *S/e4* is preventing *S/e4* mRNA accumulation in soybean seeds.

The high level of nucleotide sequence conservation observed among the putative introns of the *S/e4* pseudogene and its closest homologs (*S/e1* and *S/e5*) suggests that the point mutation that originated the premature stop codon may be a relatively recent event. Thus, it is conceivable that a functional allele of *S/e4* may still be present elsewhere in the soybean germplasm. It would be valuable to compare the temporal and spatial pattern of expression of such an allele with its closest homologs.

The predicted proteins encoded by the *S/e* gene family differ mostly by the repetition of a 20-amino-acid internal hydrophilic motif. A similar situation occurs with the homologous barley *B19* [10], cotton *D19* [17] and *Arabidopsis AtEm* [18] proteins, with some proteins having as many as four of the 20-amino acid repeats. It has been proposed for the *B19* proteins of barley that their variable number of hydrophilic repeats, their differential response to ABA and osmotic stress, and their different mRNA abundance impart different functions in the cell [10]. Our data show that in soybean the two classes of expressed *S/e* genes have different temporal

patterns of expression, with the genes encoding proteins that contain two copies of the hydrophilic motif (*S/e1* and *S/e5*) being expressed at much earlier stages of seed development than the genes containing a single copy of the motif (*S/e2* and *S/e3*). It will be interesting to determine whether the spatial patterns of expression also differ among these genes.

The hormone ABA has been implicated to different degrees as a regulatory factor in induction of the *Lea* genes [8,16,26,44,48]. *S/e* genes show only a modest response to exogenous ABA in comparison to their Group 1 *Lea* homologs in other species [10,60,61]. Also, expression of *S/e* genes, unlike its counterparts in wheat [61], barley [10], and maize [6,60], is induced by excision and culture of immature embryos, a procedure which has been shown to reduce endogenous ABA content in soybean [1]. Unlike most plant species where *Lea* gene expression has been described [43,44,53], the concentration of ABA in soybean seeds shows a single peak during the first third of seed development (from 0 to 35 DAF in our experiment) and is found at reduced levels at the time just prior to desiccation (75 DAF in our experiment) [1]. Moreover, the sensitivity of *in vitro* cultured soybean embryos to the stimulatory effect of ABA on storage protein mRNA accumulation decreases throughout seed development [5,9]. In contrast, accumulation of high levels of *S/e* mRNA does not start until 70 DAF, and reaches its maximum after 80 DAF. Thus, our results are consistent with a possible role of ABA as an environmental modulator of *S/e* expression in soybean, but not as the developmental trigger of *S/e* induction in embryos. Clearly, as ABA does not induce *S/e* mRNA accumulation in soybean seedlings or adult plants, some other type of embryo-specific developmental signal is essential for the induction of this gene family. We have yet to determine whether the functional *S/e* genes show any individual differential expression in response to ABA.

Desiccation of immature zygotic embryos in the pods induces expression of genes normally expressed in mature soybean embryos, including some Lea genes [24,50]. Under similar conditions of water deficit, we were not able to induce Sle mRNA accumulation. The induction of Sle mRNA accumulation in zygotic embryos by ABA, but not water deficit, differentiates between the response of the *S/e* gene expression to ABA and to desiccation. These results contrast with those for cotton [26], wheat [4,48], barley [10], and maize [6], in which both desiccation and osmotic stress induces Group 1 Lea mRNA accumulation in young zygotic embryos. Indeed, detailed comparisons of the responses of Lea genes to development and various environmental stimuli emphasize the distinctness of responses of embryos of different species, superimposed over the general phenomenon that these genes are expressed during late embryo development.

In vegetative organs, the expression of many of the Lea genes coding for Group 2, 3 and 4 Lea proteins is induced by ABA or water deficit [8]. This, together with the accumulation of Lea mRNAs in seeds just prior to desiccation and structural analysis of the encoded proteins, has led to the suggestion that Lea proteins may protect the mature plant under conditions of water deficit [8]. However, despite similarities in the timing of expression, each Group of Lea gene encodes a type of protein of distinct structure and potentially diverse function. The failure of vegetative organs to respond to either ABA or water deficit by Sle mRNA accumulation indicates that the Group 1 Lea proteins may not function *in situ* to protect the vegetative plant against water deficit.

Acknowledgments

The authors would like to thank Drs. Daisuke Shibata and Robert F. Whitier for providing us with a cDNA library, Dr. Richard Meagher for the soybean actin cDNA, pSAC-7, and Dr. Kayla Polzin for assistance in assembling the λ EMBL3 restriction maps. E. S. C. is grateful for his financial support provided by the Brazilian National Council for Scientific Development (CNPq).

References

1. Ackerson RC: Regulation of soybean embryogenesis by abscisic acid. *J Exp Bot* 35: 403-413 (1984).
2. Almoguera C, Jordano J: Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and *Lea* mRNAs. *Plant Mol Biol* 19: 781-792 (1992).
3. Baker J, Steele C, III LD: Sequence and characterization of 6 *Lea* proteins and their genes from cotton. *Plant Mol Biol* 11: 277-291 (1988).
4. Berge SK, Bartholomew DM, Quatrano RS: Control of the expression of the wheat genes by abscisic acid In: Ed. Goldberg R *The Molecular Basis of Plant Development*, Alan R. Liss, New York, NY (1989).
5. Bray EA, Beachy RN: Regulation by ABA of β -conglycinin expression in cultured developing soybean cotyledons. *Plant Physiol* 79: 746-750 (1985).
6. Butler WM, Cuming AC: Differential molecular responses to abscisic acid and osmotic stress in viviparous maize embryos. *Planta* 189: 47-54 (1993).
7. Chamberlin MA, Horner TH, Palmer RG: Early endosperm, embryo and ovule development in *Glycine max* (L.) Merr. *Int J Plant Sci* 155: 421-436 (1994).
8. Dure III L: The *lea* proteins of higher plants In: Ed. Verma DPS *Control of Plant Gene Expression*, 325-335. CRC Press, Inc., (1993).
9. Eisenberg AJ, Mascarenhas JP: Abscisic acid and the regulation of synthesis of specific seed proteins and their messenger RNAs during culture of soybean embryos. *Planta* 166: 505-514 (1985).

10. Espelund M, Saeboe-Larssen S, Hughes DW, Galau GA, Larsen F, Jakobsen KS: Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *Plant J* 2: 241-252 (1992).
11. Feinberg AP, Volgelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132: 6-13 (1983).
12. Finkelstein RR, DeLisle AJ, Simon AE, Crouch ML: Role of abscisic acid and restricted water uptake during embryogeny in Brassica. In: Eds. Fox JE, Jacobs M. *Molecular Biology of Plant Growth Control*, 73-84. Alan R. Liss, Inc., New York, NY (1987).
13. Frohman MA, Dush MK, Martin GR: Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene specific oligonucleotide primer. *Proc Nat Acad Sci* 85: 8998-9002 (1988).
14. Futers TS, Vaughan TJ, Sharp PJ, Cumming AC: Molecular cloning and chromosomal location of genes encoding the "Early-methionine-labelled" (Em) polypeptide of *Triticum aestivum* L. var. Chinese Spring. *Theor Appl Genet* 80: 43-48 (1990).
15. Galau GA, Hughes DW, III, Dure L: Abscisic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. *Plant Mol Biol* 7: 155-170 (1986).
16. Galau GA, Jakobsen KS, Hughes DW: The controls of late dicot embryogenesis and early germination. *Physiol Plant* 81: 280-288 (1991).
17. Galau GA, Wang HY-C, Hughes DW: Cotton *Lea4* (D19) and *LeaA2* (D132) group 1 *Lea* genes encoding water stress-related proteins containing a 20-amino acid motif. *Plant Physiol* 99: 783-788 (1992).
18. Gaubier P, Raynal M, Hull G, Huestis GM, Grellet F, Arenas C, Pages M, Delseny M: Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Mol Gen Genet* 238: 409-418 (1993).
19. Goldberg RB, Barker SJ, Perez-Grau L: Regulation of gene expression during plant embryogenesis. *Cell* 56: 149-160 (1989).
20. Goldberg RB, Paiva Gd, Yadegari R: Plant embryogenesis: zygote to seed. *Science* 266: 605-614 (1994).
21. Grandbastien MA, Berry-Lowe S, Shirley BW, Meagher RB: Two soybean ribulose-1,5-biphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences. *Plant Mol Biol* 7: 451-465 (1986).

22. Hattori T, Vasil V, Rosenkrans L, Hannah LC, McCarty DR, Vasil IK: The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev* 6: 609-618 (1992).
23. Hollung K, Espelund M, Jakobsen KS: Another *Lea* B19 gene (Group1 *Lea*) from barley containing a single 20 amino acid hydrophilic motif. *Plant Mol Biol* 25: 559-564 (1994).
24. Hsing YC, Chen Z, Chow T: Nucleotide sequences of a soybean complementary DNA encoding a 50-kilodalton late embryogenesis abundant protein. *Plant Physiol* 99: 354-355 (1992).
25. Hughes DW, Galau GA: Temporally modular gene expression during cotyledon development. *Genes Dev* 3: 358-369 (1989).
26. Hughes DW, Galau GA: Developmental and environmental induction of *Lea* and *LeaA* mRNAs and the postabscission program during embryo culture. *Plant Cell* 3: 605-618 (1991).
27. Jakobsen kS, Hughes DW, Galau GA: Simultaneous induction of postabscission and germination mRNAs in cultured dicotyledonous embryos. *Planta* 192: 384-394 (1994).
28. Jofuku KD, Goldberg RB: Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *Plant Cell* 1: 1079-1093 (1989).
29. Jofuku KD, Schipper RD, Goldberg RB: A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *Plant Cell* 1: 427-435 (1989).
30. Keim P, Diers BW, Olson TC, Shoemaker RC: RFLP mapping in soybean: an association between marker loci and variation in quantitative traits. *Genetics* 126: 735-742 (1990).
31. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181 (1987).
32. Lee JS, Verma DPS: Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization. *EMBO J* 3: 2745-2752 (1984).

33. Litts JC, Colwell GW, Chakerian RL, Quatrano RS: The nucleotide sequence of a cDNA clone encoding the wheat Em protein. *Nucleic Acids Res* 15: 3607-3618 (1987).
34. Litts JC, Colwell GW, Chakerian RL, Quatrano RS: Sequence analysis of a functional member of the Em gene family from wheat. *DNA Sequence-J DNA Sequencing and Mapping* 1: 263-274 (1991).
35. Litts JC, Erdman MB, Huang N, Karrer EE, Noueiry A, Quatrano RS, Rodriguez RL: Nucleotide sequence of the rice (*Oryza sativa*) Em protein gene (*Emp1*). *Plant Mol Biol* 19: 335-337 (1992).
36. Lopes MA, Larkins BA: Endosperm origin, development, and function. *Plant Cell* 5: 1338-1399 (1993).
37. Marcotte WR, Russel SH, Quatrano RS: Absciscic acid responsive sequences from the Em gene of wheat. *Plant Cell* 1: 969-976 (1989).
38. McCarty DR, Carson CH: The molecular genetics of seed maturation in maize. *Physiol Plant* 81: 267-272 (1991).
39. McCarty DR, Hattori T, Carson CB, Vasil V, Lazar M, Vasil IK: The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66: 895-905 (1991).
40. Murashige T, Skoog F: A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant* 15: 473-498 (1962)
41. Naito S, Hirai MY, Chino M, Komeda Y: Expression of a soybean (*Glycine max* [L.] Merr.) seed storage protein gene in transgenic *Arabidopsis thaliana* and its response to nutritional stress and to abscisic acid mutations. *Plant Physiol* 104: 497-503 (1994).
42. Nielsen NC, Dickinson CD, Cho T, Thanh VH, Scallan BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB: Characterization of the glycinin gene family in soybean. *Plant Cell* 1: 313-328 (1989).
43. Paiva R, Kriz AL: Effect of abscisic acid on embryo-specific gene expression during normal and precocious germination in normal and *viviparous* maize (*Zea mays*) embryos. *Planta* 192: 332-339 (1994).
44. Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J: Regulation of gene expression programs during *Arabidopsis* seed development: Role of the *ABI3* locus and of endogenous abscisic acid. *Plant Cell* 6: 1567-1582 (1994).

45. Perez-Grau L, Goldberg RB: Soybean seed protein genes are regulated spatially during embryogenesis. *Plant Cell* 1: 1095-1109 (1989).
46. Plant AL, Cohen A, Moses MS, Bray EA: Nucleotide sequence and spatial expression pattern of a drought- and abscisic acid-induced gene of tomato. *Plant Physiol* 97: 900-906 (1991).
47. Quatrano RS: Regulation of gene expression by abscisic acid during angiosperm embryo development. *Oxford Surveys Plant Mol Cell Biol* 3: 467-477 (1986).
48. Quatrano RS, W. R. Marcotte J, Guiltinan M: Regulation of gene expression by abscisic acid In: Ed. Verma DPS *Control of Plant Gene Expression*, 69-90. CRC Press, (1993).
49. Raynal M, Depigny D, Cooke R, Delseny M: Characterization of a radish nuclear gene expressed during late seed maturation. *Plant Physiol* 91: 829-836 (1989).
50. Rosenberg LA, Rinne RW: Protein synthesis during natural and precocious soybean seed (*Glycine max* [L.] Merr.) maturation. *Plant Physiol* 87: 474-478 (1988).
51. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467 (1977).
52. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: a laboratory manual*, Ed 2, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. (1989).
53. Skriver K, Mundy J: Gene expression in response to abscisic acid and osmotic stress. *Plant Cell* 2: 503-512 (1990).
54. Thomas TL: Gene expression during plant embryogenesis and germination. *Plant Cell* 5: 1401-1410 (1993).
55. Ulrich TU, Wurtele ES, Nikolau BJ: Sequence of EMB-1, an mRNA accumulating specifically in embryos of carrot. *Nucleic Acids Res* 18: 2826 (1990).
56. Vilardell J, Martínez-Zapater JM, Goday A, Arenas C, Pagés M: Regulation of the *rab17* gene promoter in transgenic *Arabidopsis* wild-type, ABA-deficient and ABA-insensitive mutants. *Plant Mol Biol* 24: 561-569 (1994).
57. Voelker TA, Moreno J, Chrispeels MJ: Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation. *Plant Cell* 2: 255-261 (1990).

58. Wadsworth GJ, Redinbaugh MG, Scandalios JG: A procedure for the small-scale isolation of plant RNA suitable for RNA blot analysis. *Anal. Biochem.* 172: 279-283 (1988).
59. Walling L, Drews GN, Goldberg RB: Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci USA* 83: 2123-2127 (1986).
60. Williams B, Tsang A: A maize gene expressed during embryogenesis is abscisic acid-inducible and highly conserved. *Plant Mol Biol* 16: 919-923 (1991).
61. Williamson JD, Quatrano RS: ABA-Regulation of two classes of embryo-specific sequences in mature wheat embryos. *Plant Physiol* 86: 208-215 (1988).
62. Wilson RF: Seed metabolism In: Ed. Wilcox JR *Soybeans: Improvement, Production, and Uses*, American Society of Agronomy, Inc., Madison, WI (1987).
63. Wurtele ES, Wang H, Durgerian S, Nikolau BJ, Ulrich T: Characterization of a gene expressed early in somatic embryogenesis of *Daucus carota*. *Plant Physiol* 102: 303-312 (1993).

Table I. DNA sequence identity (%) among the members of the *Sle* gene family.

	Coding Region				Intron				3'-UTR			
	<i>Sle2</i>	<i>Sle3</i>	<i>Sle4</i>	<i>Sle5</i>	<i>Sle2</i>	<i>Sle3</i>	<i>Sle4</i>	<i>Sle5</i>	<i>Sle2</i>	<i>Sle3</i>	<i>Sle4</i>	<i>Sle5</i>
<i>Sle1</i>	71.7	77.5	88.7	88.15	33.6	31.6	58.2	47.2	25.1	25.5	56.1	52.1
<i>Sle2</i>		87.2	72.8	67.7		44.0	33.3	30.0		55.7	24.0	25.0
<i>Sle3</i>			76.8	70.8			28.1	27.7			23.6	27.4
<i>Sle4</i>				89.4				65.0				88.5

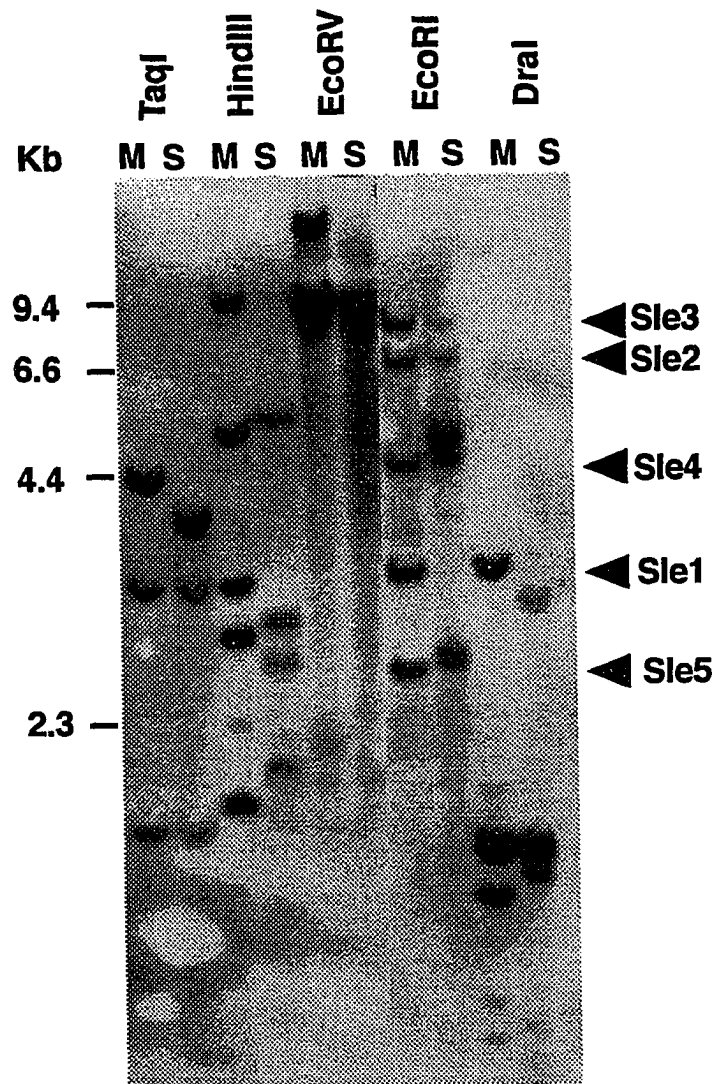


Figure 1. Southern blot analysis of *G. max* (M) and *G. soja* (S) DNA (15 μ g/lane) digested with five different restriction endonucleases. The blot was probed with [α - 32 P] labeled Sle1 cDNA. Final wash was performed in 1 X SSC, 0.1 % SDS at 50 $^{\circ}$ C. Arrowheads indicate the identity of each *G. max* EcoRI fragment visible in the autoradiograph as predicted by the restriction map of each clone.

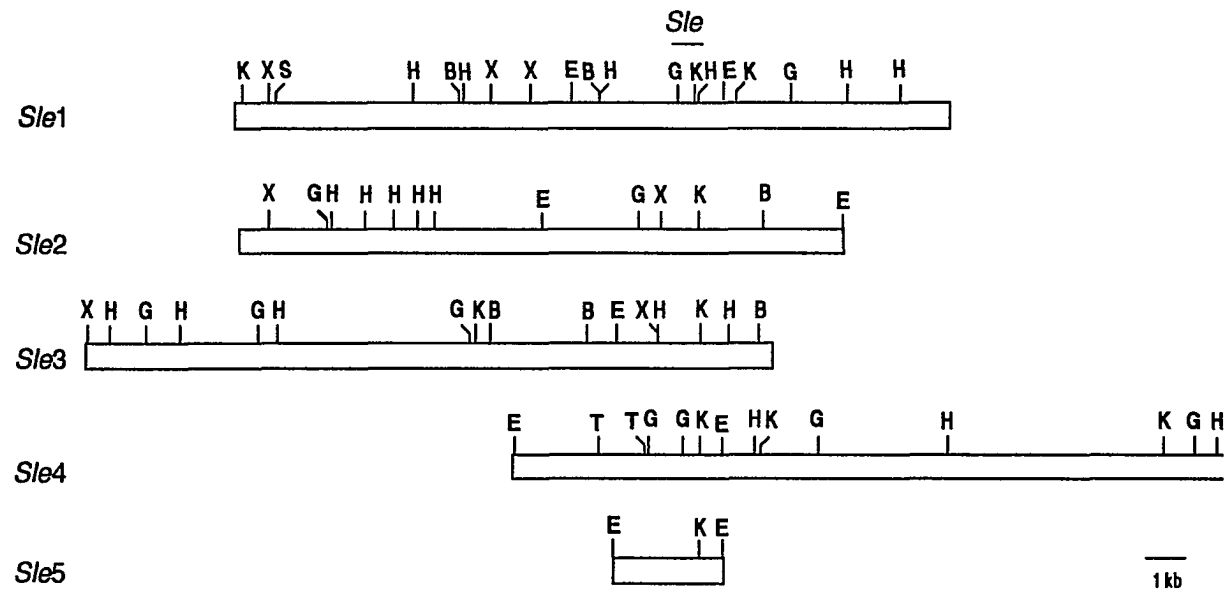
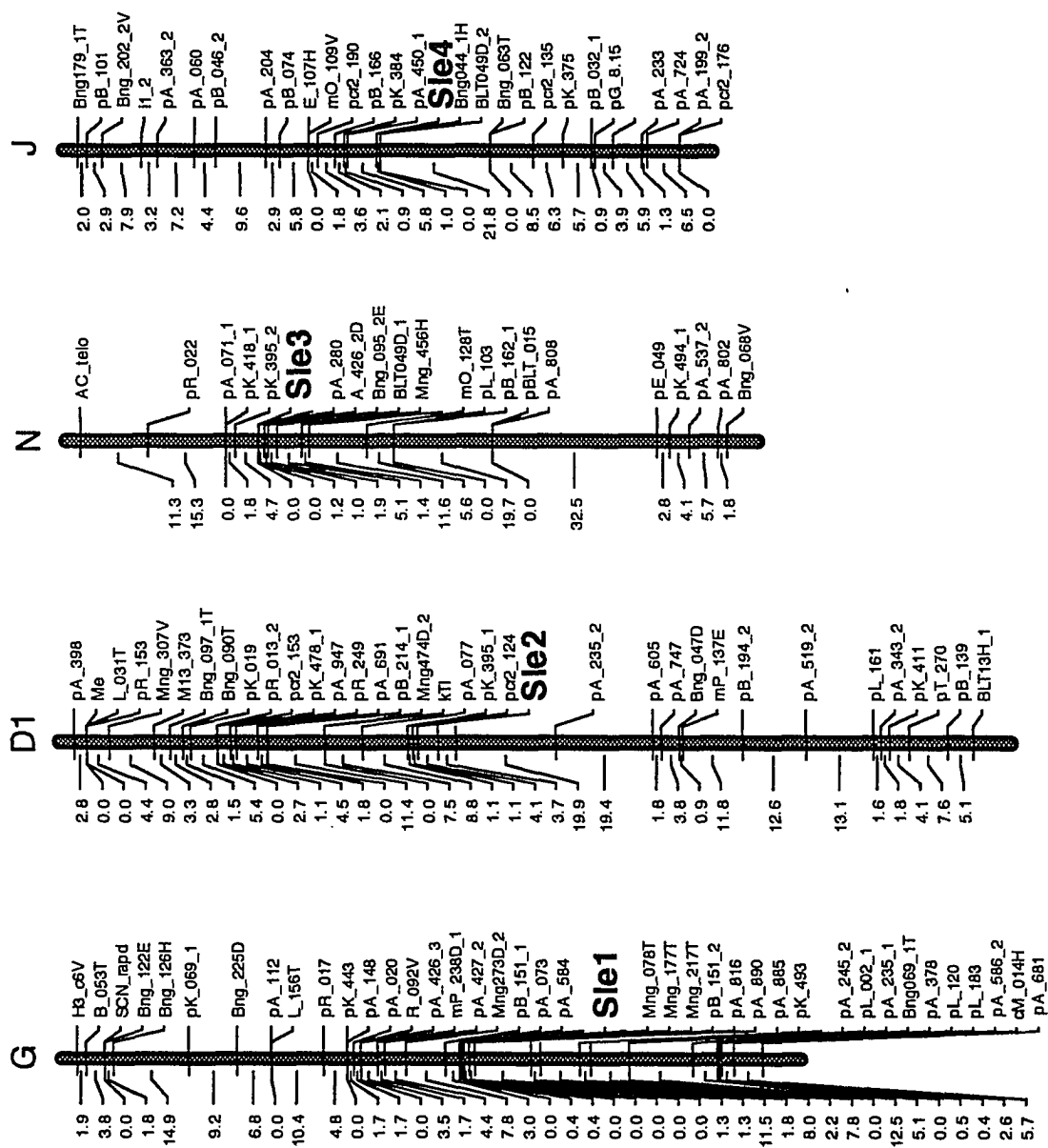


Figure 2. Restriction maps of five lambda genomic clones with six different enzymes: BamHI (B), BglII (G), EcoRI (E), HindIII (H), KpnI (K), Sall (S), XbaI (X). The fragment comprised by the two TaqI sites shown on the *Sle4* map was used to position the *Sle4* clone into the RFLP map. Not all TaqI and EcoRI sites were mapped in these clones. The region encompassed by the start and stop codons for the *Sle* protein is shown by the line on top of the figure.

Figure 3. Locations of four *S/e* loci in the soybean USDA-RFLP linkage map. Numbers to the left of each linkage group represents distances between markers, in centimorgans.



Sle1

```

1  ATGGAATCTCAGCAAGCAAACCGTGAAGAACTTGATGAGAAGGCAAGGCAAGGTGAAACT
   M E S Q Q A N R E E L D E K A R Q G E T

61  GTTGTTCCTGGAGGAACTGGTGGGAAGAGTCTTGAGGCTCAAGAACATCTTGCTGAAGgc
   V V P G G T G G K S L E A Q E H L A E G

121  atgcaatacaaccctctctcttctctttcaagaacacacatgtttcttttattattct

181  tcttaagaatcaaataataagtactaaaaagtttataataactcatctatcttgctcaatt

241  aactgacttagacccttgcatcaagaaaacacatgttaattgtcaatttttttcattgaa

301  gtaatgcattttttgtgttcacttttttaaaaaaatgagtttttatgatgtgaatgcatta

361  atgtgtagGAAGGAGCCGTGGAGGGCAGACGAGGAAGCAGCAGCTAGGGTCAGAGGGGTA
      R S R G G Q T R K Q Q L G S E G Y

421  TCATGAAATGGGGACCAAAGGAGGGCAGACAAGGAAGGAACAGATGGGGAGAGAAGGGTA
      H E M G T K G G Q T R K E Q M G R E G Y

481  CCAAGAGATGGGACGCAAAGGAGGACTCAGCACCATGGACAAGTCTGGTGGAGAACGTGC
      Q E M G R K G G L S T M D K S G G E R A

541  TGAAGAGGAAGGCATAGAAATAGATGAGTCTAAGTTTAAATTTACCTAAAAAACTAAAA
      E E E G I E I D E S K F K I T *

601  AATTGTGAGGAAGAAGTGTGTTTGTGTAGTTGTTTTAGTGTGTGTGTGTGTGTTTTTTGA

661  TGTAGTAGTGAATATAGCTAGGAACTAGCTAGTTCTATAAAATATATGTTTTAAAGCTT

721  TAGTGCTACTGTTTGTGTTAGGAATGGTAGGTATTTTAGGTTAAGGAACCTTAGGTAGTCTT

781  TATCTTAGGGGTATGTTTATCATGGTGTGTAGGAAGTGTAACTATCAGTGGAGACTC

841  TTGTACCCCTTATAATGCATGTCCAATTATCTTTAAACATTGAGTTCCATCAAGTTTTTA

901  AGAGTGGCCTTATGACAGTTGTCTTCAGAATATAATATAAGGAGCAAGGAGGATAAATAG

961  AAATAAAAAAGAAATTGAAGCATGTTTTATGTCGCGTATAAATTTTATCAATGTGATAAA

1021  TGAACGCTGAATGACAAGTATCATGAAAGATAAAAACAATGAGTGGTTCACACAATTGAC

1081  CCTTGTTTTGTGGTGTGTTATGAATTGTATGAATATGAATATATGATGAATGAATTCTGCG

```

Figure 4. DNA sequence of the five *Sle* genes. Intron sequences are shown in lowercase. Sequences of the deduced polypeptides are shown underneath each DNA sequence. The 20-amino acid hydrophilic motif appear underlined. The start and the end of the *Sle1* cDNA is shown with an arrowhead above the *Sle1* sequence. Except for *Sle1*, intron positions and deduced polypeptides were predicted based on conservation among all the *Sle* homologs.

1141 ATGAACGGTGAATGGTGGAAACGGATGAAACGGGAGGGAATAATTGACAATAGCTTTAAAT
 1201 TATTTTTTTTATAATCTCAAATCATAGAATGATAAATAAGGGTCCGTGGTTCAGGGAAT
 1261 GTATGCTCTACTCTTTGTACAATTCATTGTCCTTTAAACCTTCTTAAAAAGAAACCTCT
 1321 TCTCACCTTAAATAATCAAACTTGAATAAGTGGATAAGAGTAGAGAACAAGAAAAACA
 1381 AGGGGAAAAAACCAAACCTAGAAAGCGCAATACTGGACAAGCCTTGGTCATACTCCTTGT
 1441 ATGTAGGGTGGCCAACAAAAGCTATGTCACAAATCTTGAATGCTGGTACC

Slc2

1 ATGGCATCTCGTCAAAACAACAAGCAAGAGCTTGATGAAAGAGCAAGGCAAGGAGAGACT
 M A S R Q N N K Q E L D E R A R Q G E T
 61 GTTGTCCCGGTGGAACTGGTGGCAGAGACCTGGAGGCTCAGCAACACCTTGCTGAAggt
 V V P G G T G G R D L E A Q Q H L A E G
 121 ttataatcaattttatatatatatatatggaacattaattttaattttttcgtatatt
 181 ttttttataattgtgttgcatatccatacaattattcatcataacacttacgttacattt
 241 ctctaattttttacttccttcaaattagttttattcatcatctatcatatctgaccatc
 301 agatttttgtctatattattcatcacgtaattttattcatcatcttcattgaaatcattat
 361 cattatctatatcatactagttccatcacatttttatogattttttctttttactttcttt
 421 tccttggttttttttttttttttttgagtatagacactttcattaggcacaacaacaacaag
 481 gttttttttcacaaatatctcaatcggtatttttatctcaagtgtttgatgaagttatttt
 541 tgtgcatgatattttttttcttctttttttctatttgactttggattttggggtttatttt
 601 attgttgattgtgcatatatataaaaagGAAGGAGGAAGGGAGGGCAAAACAGGAAGGAAC
 R R K G G Q T R K E Q
 661 AGCTGGGTACAGAAGGGTACCAGGAAATGGGACGCAAAGGAGGGTTGAGCACTGTGGAGA
L G T E G Y Q E M G R K G G L S T V E K
 721 AATCAGGAGAAGAACGTGCTCAAGAGGAAGGCATTGGCATCGATGAGTCCAAGTTCAGGA
 S G E E R A Q E E G I G I D E S K F R T
 781 CTGGTAATAACAAGAACCAGAATCAGAACGAAGACCAGGATAAGTGATAATGTAACACCC
 G N N K N Q N Q N E D Q D K * *

Figure 4 cont.

821 TTCACTAGCATGTGTAGTGTGTTCTTATGCTAGCTTAAATGTGAGTGTCAAGTGTGTTT
 881 GAAGTAGGAAAAATAATAGGTTATAAAAAATAATGTTTAGGCAATAGTTATATAGGTCAA
 941 GTGTCATGTCATGTGTGTCGTGAAATAATTAGCTAGCATGTTTAGGTTGTTTTAGGGTTT
 1001 CTTTTGCATGTGTGTTTCTCTTATGTGTGTAAGTGATCAGTGAGGCATGGCTCTATAGCT
 1061 ATCCAGAAATTAGCACAGCAAGCATGTGCCTCCATTTCACTAATGTGTATGTATCTATG
 1121 GTGTGACCACGTTCAATACTAGAATGTAATGGTCTCAACTCTTATGTTAACTTATCGTGG
 1181 CACAAACAATTACGTACGTACTCAACTTAATCCCTTTTTTTTTTTTTGGTCTTGCATCAA
 1241 TGCTTGTTAGCAGCAATACCTGAAGTTGCTCTAGGTTAACATCTTTCTTTTTTTCTTT
 1301 AAATTAAATATCTTTTCAAACCTGTATTACGAATATTATCACGAAATCCATGCTGTTGT
 1361 TGTCTAATTACGAATACTATCAATTGTCTAATTACGATAATTATCAACAAAAAAGTATA
 1421 ATGTTGTCTAATGACTTCGGTAAGTGAACGTCGTCGTCTTGAAACTCATACTGATTTTCT
 1481 GTCGTCTTGAAACTCATACTGATTTTCTGTCTACGATTAAAAATCTAAAAAATGGAAAA
 1521 TTACAATTGCTAGATGTAAATTTGTAATGCTTGCATATACTACAAATAAAACGTCTTA
 1581 TATCCACATGGTTTTGCTACAAAATGCAGCAGCTCTAGTTTTTGGTCGTCTTAATGACTA
 1641 TTAAACCTCAACATATTAGCCCTCCATATTTTGATAAATATTTGGAAAAAGAAAGATCAT
 1301 GCGAAATTTACGTGGACTTAAGAAATTTCCATGCATGAGCATCTATTCAATTCTCTAAGA
 1361 TTAAGTTGAACTGATTTTAATATGTAGACGTTAAAATCTTTTTTAAACAATTATGAGA
 1421 ATTTTCTTAAAACTAGTTTGT

Slø3

1 ATGGCATCTCATCAACAAAACAAGCAAGAGCTTGATGAAAGGGCAAGGCAAGGAGAGACT
 M A S H Q Q N K Q E L D E R A R Q G E T
 61 GTTGTTCGGGTGGCACTGGTGGCAAGAGCCTTGAGGCCAGCAACATCTTGCTGAAGgt
 V V P G G T G G K S L E A Q Q H L A E G
 121 ttattatcaatttatagaacactaattattattggtttccatactttttactttttttttt
 181 tttatatttgtagatgaatctctcttatttttaaaattttattttcaaagtcttatt
 241 attcatcatgacactttacatatttcttctcgttctttttactttcttttcagattatta

Figure 4 cont.

301 ttatacatcacatcacttaatcattttcaaattattatacatcacattttattcatcatttt
 361 cattgaaatcacattttattcatcatatttggccatcaaatttttatctatattattcttt
 421 tcattaaaatcattattacctatgtcattcatcattttctttgaaatcactgttatctat
 481 catatttgtccatcacatttttggcgatattattcattacatttggccatcattttcca
 521 ttgaaacctatcatatttggtccttcacatatttttcaattttctttttttatctttca
 581 ttttcctgaaatcatatcatcattttttattttactttcttatcctttttaagttatgga
 641 cactgtcattagaggcgtaaaaaagttccatcttttccataaatatctcagatcactatt
 701 aatttcttctcaagtgaatgattgatgaagtcatttttgtgtttggtattttgtttgat
 761 attccataaatatctcaaccattatttttctctcgaagtgcgattgatgaattatttta
 821 gtgtgtttcggtattctgttgataacaataactgacttttggttttggggtttgttatt
 881 tttgggttattgtaatccatattaactagGAAGGAGCAGGGGAGGGAAAAACAAGGAAGGA
 R S R G G K T R K E
 941 GCAGCTGGGGACAGAAGGGTACCATGAAATGGGACGCAAAGGTGGGTGAGCACTATGGA
 Q L G T E G Y H E M G R K G G L S T M D
 1001 CAAGTCAGGAGAAGAACGTGCTCAAGAGGAAGCCATTGACATCGATGAGTCCAAGTTCAG
 K S G E E R A Q E E A I D I D E S K F R
 1061 GACTGGTAATAACCAAGACAAGAACCAGAACAAGTGATCATGCAACACCCTTCACTATCA
 T G N N Q D K N Q N K *
 1121 TATGTAGTGTGTTCTTATGCTAGCTTAAATGCTAGTGTCTTAGTGTTTTAGGAAAGAATA
 1181 GGTATATAAAAATGGTATAGGTAGCAGCTAGGTCAAAGTGTGATGTCATTGTGTCGGTGAA
 1241 ATTATTAGCTAGCATGTTTAGTGTTTTAGGGTTTCTTTTGCATGTGTGTTTCTCTTATGG
 1301 GGTGTAAGTATGAGTGAGGCATGGCTCTACTATAGCTATCTAAAAAATAATGCAAGCTT
 1361 TTTTTTATCTCTTGTTTCGTGTAATATATTTATCGTGCCTCCATTTCACTATTATGTATGT
 1421 ATCTATGGAGACCACGTTCAATACTAGAAGTAATGTTCTTAACTATGTTAACTTATCGTG
 1481 GCAAAAAACCTATTACGTAATTCAGATCAACTTAATTCCTTTTCAATTTTATTTTTTAA
 1521 TTTTTTGACTTGATCGATGTCCAGCAACACTGAAGTTCTCTCTAGGTTTAAATCTTTC
 1581 TTCTTTTCCTTAGTAGAATTTTTTTCGTCGGAAAACAAATATCTTCCAAAACGAATATT
 1641 ATCACAAAAATACATGCTGTTCTCTAATCACGAATAATATTATAAAGCAAATCATGCCAT

Figure 4 cont.

1701 TGTCGAATGACTTTGGTAAGGGAACGTCGTCGTCCTTGAAGCAACTAAATTAATGGAGTGA
 1761 AAAAAATATTAGAAATGATAATAAATACTGATTTGCTGTCTATGAATAAATCTCTAAATT
 1821 TTTAAAAAATCACAACCTCCTACATGTAAATTTTGTAATGCTTGCATACAAATAAGCGTCT
 1881 TTATTTTGTCTCCATGCTTTTACTACAAAATTAAGCAGCAGCTCTAGTTTG

Sle4

1 ATGGAATCTCAGCAAGCAAACCGTGAAGAACTTCATGAGAGGGCAAGGTAAGGTGAAACT
 M E S Q Q A N R E E L H E R A R * G E T
 61 GTGGTTCCTGGTGGAAATGGCGGGAAGAGTCTTGAGGCTCAAGAACATCTTGCTGAAAgc
 V V P G G N G G K S L E A Q E H L A E G
 121 atgcaaTacaaccccttcctcttcctctttcaataacacatgttttttttttaatatgta
 181 ggaatcagatacaggcgaatacaggtataaggggtttacgtttataataactcgcgtatc
 221 ttgtttaaccaattgaactaataaatccctttatcaagagcatacatgttaattgtcaat
 281 tttttcattgaagtatttgtggttactttataaaaaatgagtttttatgatgtgaatgaa
 321 tgtgtagGAAGGAACCGTGGAGGGCAGACGAGGAAGGAGCAGATGGGGAGAGAAGGGTAC
 R N R G G Q T R K E Q M G R E G Y
 381 CAAGAGATGGGACCCAAAGGAGCCCTCAGCACCATGGACAAGTCTGGTGTAGAGCGTGCT
 Q E M G P K G A L S T M D K S G V E R A
 421 GAAGAGGAAGGCATAGACATCGATGAGTCTAAGTTTAAGAATAAATAAAGCATGTTTGAG
 E E E G I D I D E S K F K N K *
 481 CAAGTAGTGATTTGTTGTAGTTGCTTTAGCTAATATGTGTGTTTTGATGTAGTAGTGAAT
 521 ATGTGTAGGAGACTAGTACTACAAAATATATGTTTTGATGCTCCTGTTTGGTAGGAATGG
 581 TTGGTATTTTAGGTTTAGGAACCTTAGGTAGTCTTTATCGATCTTAGGGCATGATATGGA
 641 TCATGCTTATCATGATGGTGTGTAGGAAGTGTAAACTATCACTGGAGACTCTTTTACT
 701 ACTTATAATACATGTTCTCTTATTTTTAAACGTTTGAGTTCCATCAAGCAGCTTTTAAGA
 761 AGTGGTCTTAAGAAAGTTGCATTTAGAATATAAGGAACAAGGAGGGTAAAGAGAAGCTAA
 821 TATGAGATTGAAAGGACGGATACAGATTACACAATCAAGGCATGTTTTATGTAGTGTATA
 881 AATTTGATCAATGTGAAAGAATGGACATCATGCAATTGAATCAAGTTTTGAGTTGTTTAT

Figure 4 cont.

941 AAGTTGTATAAATACATGTTAAATGAATTCTGTCATGAACTCAATAACGTGAGTGGTGGG
 1001 ACGGATGAAGAGTCGAGAACAATAACTAAAGTTTCAAGTTTTCATCTCAAATCAATGACG
 1061 GGCCAAACATAAGGGTCCCAGATTTCGAGGGAACAATAACATGTCTGTGTCTACACTTTGT
 1121 ACAATTTATTGTCTTTTAAAGCTTCCTAAAAAGAAACCCCTTCTCATCTAAAATAATCC
 1181 AAACCTGAACAAGTGGAAACGAGGAGAACAAAGGAAAAAAAAAAAAACACAATACTGGACA
 1221 AGCCTAGGTCATACTCTCCTTGAATGTAGGGTGGCCAAACAAAAGCTATGCCACAAATCT
 1281 TGAATGCAGGTACC

Sle5

1 gattatatatattgtcatgtctacttttatatgtcaaacataataacttgagattgttatg
 61 tgtccaaagtttgactctgatactttattgatttcgttttatgcaattatagaaaaatgt
 121 tcaacgatcttgagatgaaaatagtccttcaaattttccattagatagtcaaatagtcatt
 181 tgatgatggaatcttcaaactggtgtcccgatatagtctgaagacaatattcatggattgt
 241 atatctttcaataagatgaattatcattattcaatctctatataacttcagtgacattcgc
 301 tctatgcccgaatcgctatgaattgattgtctaatttattttttaccggttaaaactatac
 361 gattgttgaaatcatatttccttaacaatcggtttgacagggttaatacccaatttatgcgc
 421 accaattgtgcaagcacaaatgaactagttggttgatctgtttccatttcaatcttaact
 481 atttaatatgtcttttcttaaacataaaatttcaaattcaatggttatataataataggg
 541 aatgcatgttgattagcttttctagtcacgtatttgagcagcttgctgctatatatcatc
 601 acaatttaattgcacttgtagcgattgcaagtacacttggtgcatttatattcctaaatca
 661 ctaaaatctatctgttggttttacaaattatatttcgtataactcgatatatatgtaggaat
 721 cagatacaggcgaatacaggtattaagggataaggggttacggtttataataacactcgc
 781 ctatcttggttaaccaattgaactaaattcctttatcaagaacatacatgttaattgtca
 841 atttttttcattgaagtttgggtgtttacttttataaaaaatgattgtttaaccaattga
 901 actaaattcctttatcaagaacatacatgttaattgtcaatttttttcattgaagtattt
 961 gtgtttactttataaaaaatgagtttttatgatgtgaatgaatttgtagGAAGGAACCGT
R N R

Figure 4 cont.

1021 GGAGGGCAGACGAGGAAACAGCAGCTAGGGTCAGAGGGGTATCATGAGATGGGAACCAA
G G Q T R K Q Q L G S E G Y H E M G T K
 1081 GGAGGGCAGACGAGGAAGGAGCAGATGGGGAGAGAAGGGTACCAAGAAATGGGACGCAA
G G Q T R K E Q M G R E G Y Q E M G R K
 1141 GGAGGCCCTCAGCACCATGGACAAGTCTCGTGTAGAGCGTGCTGAAGAAGAAGGCATAGAC
 G G L S T M D K S R V E R A E E E G I D
 1201 ATCGATGAGTCTAAGTTTAAAGAATAAATAAAGCATGTTTGAGCAAGTAGTGTTATGTTGT
 I D E S K F K N K *
 1261 TGCTTTAGCAAATATGTGTTTCGATGTAGTAGTGAATATGTGTAGGAGACTAGTACTACA
 1321 AAATATATGTTTTGATGCTCCTGTTTGGTAGGAATGGGTAGGTATTTAATTTAGGTTTAG
 1381 GAACCTTAGGTAGTCTTTATCTTAGGGCAAGAAATGGATCATGCTTATCATGATGGTGTT
 1441 GTAGGAAGTGTTAAACTATCACTGGAGACTCTTTTACTACTTATAATACATGTTCTCTTA
 1501 TTTTAAACGTTTGAGTTCCATCAAGCAGCTTTTAAGAGTGGTCTTAAGGAAGTTGAATT
 1561 TAGAATATAAGGAACAAGGAGGGTAGAGAGAACTAATGAGATTGAAAGGCCGGATACAT
 1621 ATTACACAATCAAGGCATGTTTTAGTAGTGATAAATTTGATCAATGTGGAAGAATGGAC
 1681 ATCATGCAATAAAATCAAGTTTTGTGTTGTTTGTAAGTTGTGTAAATACATGTTAAATGA
 1721 ATTC

Figure 4 cont.

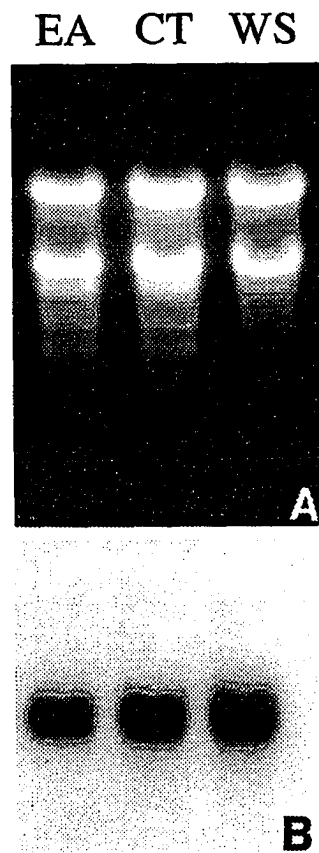


Figure 5. Sle mRNA accumulation in axis and cotyledons of seeds. Northern blot of total RNA (15 μ g/lane) isolated from mature dry seeds dissected into the embryo axis (EA) and the cotyledons (CT). Whole seeds (WS) were used as a control. Final wash was done at 2 X SSC, 0.1 % SDS at 50 °C. (A) Ethidium bromide stained gel before transfer. (B) Autoradiograph of membrane probed with [α - 32 P] labeled Sle1 cDNA.

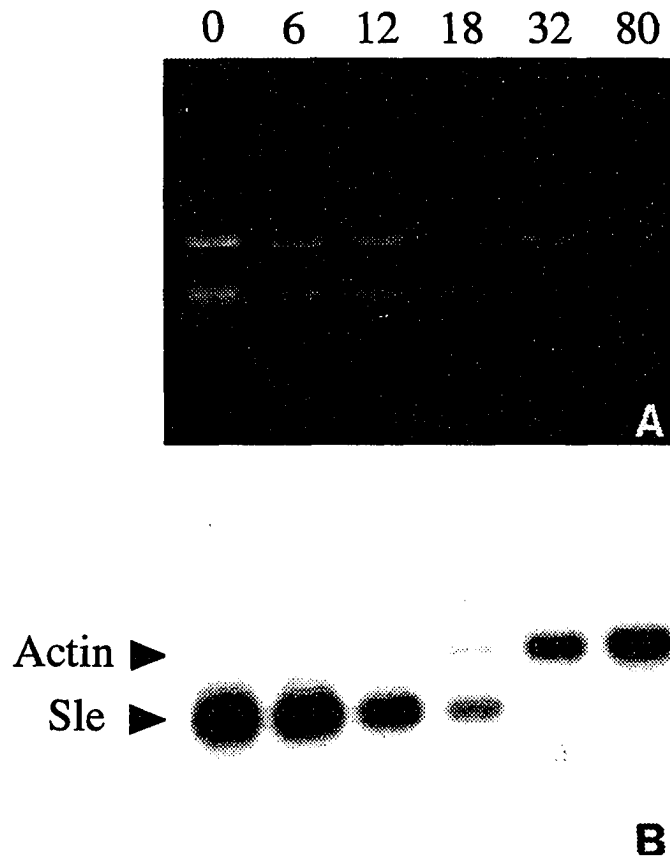


Figure 6. Sle mRNA accumulation during germination. Northern blot of total RNA (10 μ g/lane) isolated from seeds after 0 (dry seeds), 6, 12, 18, 32, and 80 h of imbibition in water. Final wash was done at 2 X SSC, 0.1 % SDS at 50 $^{\circ}$ C. (A) Ethidium bromide stained gel before transfer. (B) Autoradiograph of membrane probed simultaneously with [α - 32 P] labeled Sle1 cDNA and actin cDNA.

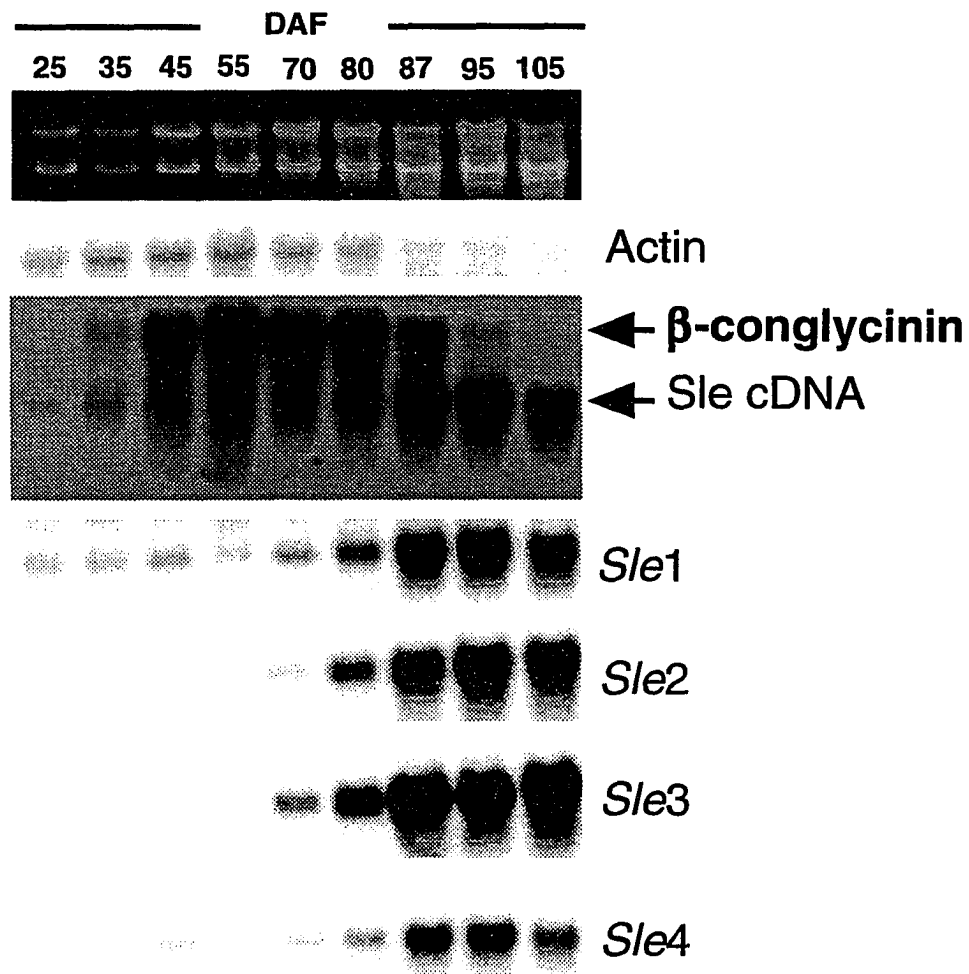
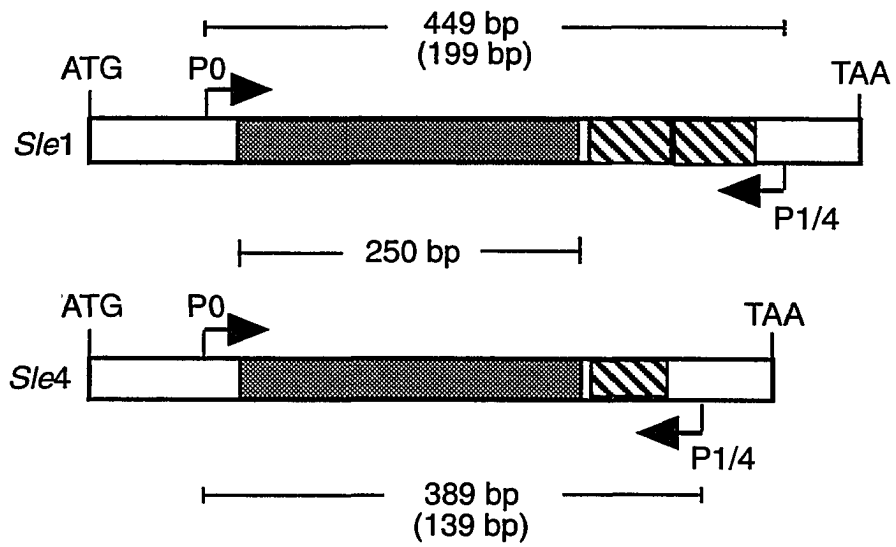
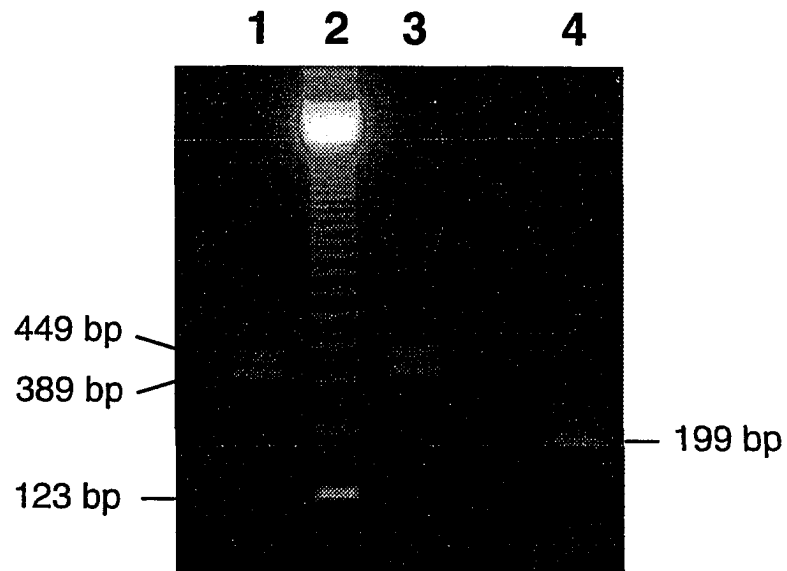


Figure 7. Developmental profile of *Sle* mRNA accumulation during soybean seed development. Northern blot of total RNA isolated from developing seeds. Seeds at 25 DAF were already at the cotyledonary stage; seeds at 105 DAF were mature and dry. The average fresh weights (mg) of seeds at each DAF were: 23 (25 DAF), 58, 120, 197, 340, 390, 400, 360, and 133 (105 DAF), respectively. All lanes were loaded with 15 μ g of total RNA. A picture of the ethidium bromide stained gel is shown at the top. The same membrane was successively probed with gene-specific probes (see methods) in the following order: Actin, *Sle2*, *Sle3*, *Sle1*, *Sle5*, *Sle* cDNA, and β -conglycinin. Membrane was stripped after hybridization with each *Sle*-derived probe. Final washing conditions for the probes were: 0.1 X SSC, 0.1% SDS at 65 $^{\circ}$ C for Actin, *Sle1*, and *Sle5*; 2 X SSC, 0.1 % SDS at 50 $^{\circ}$ C for *Sle* cDNA and β -conglycinin and; 5 X SSC, at 50 $^{\circ}$ C for *Sle2* and *Sle3* oligonucleotide probes.

Figure 8. RT-PCR analysis of expression of *Sle1* and *Sle4* loci. (A) Schematic representation of the *Sle1* and *Sle4* coding regions (blank boxes) and introns (shaded boxes). The arrows denote the location of binding sites for the primers (P0 and P1/4) used to amplify both loci. The 20-amino acid hydrophilic motif is represented by a hatched box. Predicted size of PCR products are indicated; numbers in parentheses are the expected size of fragments after splicing of the intron from each locus. (B) Ethidium bromide stained (1.3 %) agarose gel showing the results after the PCR amplification using either soybean genomic DNA (lane 1 from cv. Williams; lane 3 from cv. Williams 82) or first strand cDNA from soybean (cv. Williams) developing seed (lane 4) as template for the reactions. The 123 bp ladder appears in lane 2.

A**B**

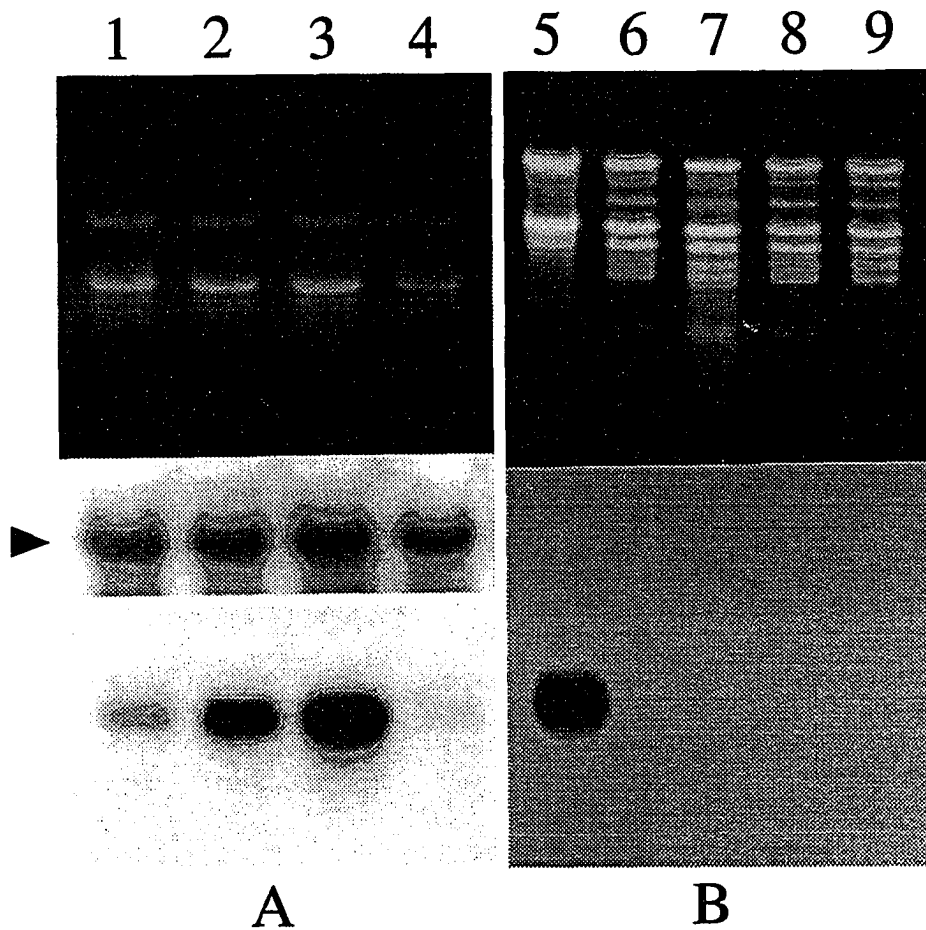


Figure 9. Effect of ABA and water deficit on the accumulation of *Sle* mRNA. Top panels: Ethidium bromide stained gels before transfer. Bottom panels: Autoradiographs of membranes probed with [α - 32 P] labeled *Sle* cDNA. Panel A also was probed with [α - 32 P] labeled actin cDNA (indicated by arrowhead). (A) Total RNA (15 μ g/lane) isolated from immature zygotic embryos. Embryos were dissected from pods 30 DAF and cultured in MS medium supplemented with or without ABA for a 24 h period, or subjected to water deficit. Lane 1, embryos at the time of dissection from pods; lane 2, embryos cultured in MS media without ABA; lane 3, embryos cultured in MS + 100 μ M ABA; lane 4, embryos left in the excised pods and allowed to dry for 48 h. (B) Total RNA (15 μ g/lane) isolated from leaves of five-week-old plants treated as indicated in methods. Lane 5, dry seed control; lane 6, plants were sprayed with 100 μ M ABA; lane 7, plants were sprayed with water; lane 8, unsprayed plants; lane 9, plants were subjected to water deficit treatment.

CHAPTER 2: PROMOTER ANALYSES OF A GROUP 1 LEA GENE FROM SOYBEAN (Glycine max (L.) Merr.) IN TRANSGENIC TOBACCO

A paper to be submitted to The Plant Journal

Éberson S. Calvo, Randy C. Shoemaker, Eve S. Wurtele

Abstract

We have previously cloned the *S/e* gene family from soybean and shown that *S/e1* is less ABA-dependent than most homologs from other plant species. Here we report the nucleotide sequence of the promoter regions for *S/e1* to 4. The *S/e1* and *S/e4* promoters each have two conserved SphI box motifs (CATGCATG) in their proximal promoter regions, one of them flanked by two CCACAT inverted repeats. Two hexameric G-box core motifs (CACGTG) are also found at the distal *S/e1* promoter region. *S/e2* and *S/e3* promoters contain only variants (CAATGCATG) of the SphI box and are devoid of the CCACAT repeats, but as all known Group 1 Lea genes, they contain the G-box core at the proximal promoter region. A transcriptional fusion of a 1823 bp *S/e1* promoter fragment was able to drive GUS (β -glucuronidase) gene expression in both the embryo and the endosperm of transgenic tobacco seeds. Several mutant versions of the *S/e1* promoter were also analyzed. The results show that neither the two CACGTG hexameric core motifs located at positions -1216 and -1248 bp nor a third ACGT-containing motif at -89 bp constitute a major site for *S/e1* promoter function in transgenic tobacco seeds. Rather, the SphI box at position -106 bp and its flanking CCACAT repeats play a major positive role in this response. This is the first report

of the involvement of these sequences in expression of late embryogenesis abundant (Lea) genes in plants.

Introduction

Analyses of the promoter region of the wheat *Em* and rice *Rab16A* late embryogenesis abundant (Lea) genes led to the identification of the first consensus DNA sequence involved in a hormonal response in plant tissues, the ABA-responsive element (ABRE; CACGTGGC and CCGTACGTGGC in wheat and rice, respectively) (Guiltinan et al., 1990; Mundy et al., 1990). This finding provided strong evidence for the existence of ABA-mediated transcriptional activation of Lea genes during seed development or in vegetative tissues subjected to water stress, two physiological processes mediated by the hormone ABA (Quatrano, 1986; Skriver and Mundy, 1990; Hetherington and Quatrano, 1991).

The ABRE shares a common ACGT core motif with several other cis-acting elements such as the G-box involved in the response to light and anaerobiosis, and the hex-1 elements involved in cell-cycle dependent gene expression (see Foster et al., 1994 for a review). These ACGT-containing motifs, including ABRE, are the binding site for an array of proteins belonging to the leucine zipper (bZIP) family of transcription factors (Foster et al., 1994). Therefore, presence of the ABRE does not necessarily imply ABA or water responsiveness. Sequences quite different from the ABRE also can confer ABA response (Thomas, 1993). This picture is further complicated by the fact that both the ABA and the water stress-response conferred by ABRE in embryos are likely to be transduced by pathway(s) different from those mediating these same responses in vegetative tissues (Finkelstein, 1993; Vilardell et al., 1994). Furthermore, the role of ABA as a major

factor regulating *Lea* gene expression in embryo development has been questioned for at least some dicot species, including soybean (Galau et al., 1991; Calvo et al., submitted). Recent experiments with viviparous mutants in *Arabidopsis* clearly show that the temporal pattern of expression of *Lea* genes may be determined by factors other than endogenous ABA (Parcy et al., 1994). In light of the emerging complexity of *Lea* gene regulation, it is imperative to investigate the promoter region of different *Lea* genes as a means to characterize cis-acting elements involved in transcriptional regulation during embryo development.

In addition to the ABRE, the promoters of all *Lea* genes contain several other ACGT core motifs (Foster et al., 1994) but their roles in *Lea* gene expression have not been reported. Moreover, some Group 1 *Lea* genes such as the wheat *Em* (Litts et al., 1991) and the rice *Emp-1* (Litts et al., 1992) have two copies of the SphI box (Hattori et al., 1992) or RY repeats characteristic of promoters from seed-specific genes (Dickinson et al., 1988). In legume storage protein genes, the SphI boxes were found to act as seed-specific enhancers (Bäumlein et al., 1992; Chamberland et al., 1992; Lelievre et al., 1992). More recently, it has been shown that they can confer a silencing effect on the promoter of a seed-specific non-storage protein gene (Fiedler et al., 1993). Studies also show that the SphI box is essential for the regulation of VP1 protein and ABA responses in the maize *C1* promoter (Hattori et al., 1992). Nonetheless, the role of the SphI box in *Lea* gene expression is still unknown.

We have cloned the entire *Sle* gene family, the *Em*-like gene family, from soybean (Calvo et al., submitted). Unlike *Em*, exogenously applied ABA is not necessary to induce *Sle* 1 expression in *in vitro* cultured embryos and it confers only a modest response when added into the culture media. Our long-term objective is to dissect the pathway(s) through which defined spatial and temporal

patterns of *S/e* gene expression are established during soybean seed development. In this paper we report the sequence of the promoter region of four *S/e* genes, and examine the role of putative regulatory motifs in the quantitative response conferred to a β -Glucuronidase (GUS) reporter gene in transgenic tobacco plants in the context of the *Lea S/e1* promoter.

Experimental Procedures

DNA Sequence and DNA Comparisons

Details of the isolation of the *S/e* genomic clones are presented elsewhere (Calvo et al., submitted). The 5'-flanking sequence of each *S/e* gene was subcloned into either pBluescriptKSII (Stratagene) or pGEM7Zf (Promega) plasmids. Unidirectional deletions were generated with the Erase-a-Base kit from Promega. Manual sequencing (Sanger et al., 1977) was performed using a Sequenase 2.0 kit from USB. DNA sequence alignment was performed with the Mac DNAsis sequence analysis software (Hitachi) using default settings in the Higgins-Sharp mode.

Construction of Plasmids

A 2320 bp fragment of the *S/e* 1 locus (Calvo et al., submitted) was subcloned as a HindIII / KpnI fragment into pGEM7Zf⁺ (Promega). This plasmid was digested with EcoRI and BglII, and into which a BamHI/EcoR I fragment from pBI221 (Clontech) containing the GUS coding region plus the nopaline synthase (NOS) terminator was ligated. The resulting construct, p2618GUS, contains the

GUS coding sequence under the control of a 1823 bp promoter fragment from *S/e1* (Figure 3). Plasmid p Δ HincIIIGUS was obtained by digesting p2618GUS with HincII followed by gel purification and religation with the vector fragment. Plasmid p Δ PmlIGUS was derived from p2618GUS after digestion with PmlI restriction enzyme and followed by treatment with T4 polymerase, gel purification and religation with the vector fragment. After transformation into *E. coli*, bacterial colonies carrying plasmids which had lost the PmlI (CACGTG) site were identified by their inability to be cleaved by PmlI restriction enzyme. The extent of the deletion was verified by sequence analysis and is shown in Figure 3 as dashes. To obtain the p Δ SphI construct the 1823 bp *S/e1* promoter fragment of p2618GUS was initially subcloned into pBSKS⁺ as a HindIII/SmaI fragment. The resulting plasmid, pBS2618, was digested with SphI restriction enzyme, treated with T4 DNA polymerase, and religated with T4 DNA ligase. A plasmid lacking the SphI site was initially identified by SphI restriction digestion. Partial sequencing confirmed the deletion of the CATG tetramer. The promoter fragment lacking the SphI site was then transferred back to the p2618GUS backbone as a HindIII/SmaI fragment. Finally, the promoter, GUS coding region, and NOS terminator from p2618GUS and its derivatives were transferred to the binary vector pBI121 (Clontech) backbone as a HindIII/EcoRI fragment.

PCR Mutagenesis

Plasmids pMP1-MP2GUS and pMP3GUS were created through PCR-assisted mutagenesis (Landt et al., 1990). Sequences of the primers used and their positions on the *S/e1* gene are presented in Table I. All PCR reactions used the plasmid pBS2618 as the template and were conducted in a 50 μ l final volume

using 1 U of Taq DNA polymerase and reaction buffer purchased from Promega. Primers, dNTPs, and MgCl₂ were used in the reactions at 0.2 μM, 200 μM, and 2 mM final concentrations, respectively. Cycling conditions were as follows: 45 sec at 93 °C (denaturing step); 2 min at 60 °C (annealing step); and 2 min at 72 °C (extension step) with a total of 30 cycles. Sequence analysis with the M13 reverse primer confirmed the introduction of the desired mutation as well as the absence of other mutations downstream of the mutation site. The resulting plasmids were named pMP1 to 3, for primers 1 to 3, respectively. To correct for the possible introduction of any undesired mutations upstream from the mutation site we took the following approaches. For plasmids pMP2 and pMP3, the entire region upstream of the mutation site was replaced by the original 1650 bp HindIII/SphI fragment from pBS2618. For plasmid pMP1, only the 1002 bp HindIII/AccI fragment was replaced by the corresponding fragment from pBS2618. The rest of the promoter (648 bp from the AccI to the SphI site) was sequenced and found to be devoid of any mutation other than the mutations introduced by the MP1 primer. Plasmid pMP1-MP2 was obtained by replacing the HindIII/SphI fragment of plasmid MP2 with the corresponding fragment from plasmid pMP1. Each promoter was cloned into p2618GUS as a HindIII/SmaI fragment, thus creating plasmids pMP1-MP2GUS and pMP3GUS which have the GUS coding region and NOS terminator under the control of the mutated promoters. These plasmids were finally transferred to the pBI121 (Clontech) backbone as a HindIII/EcoRI fragment. All plasmids (including those described in the previous section) were mobilized to *Agrobacterium* strain LBA4404 by direct transformation after freezing with liquid nitrogen, and immediately followed by thawing at 37 °C for 5 min. Integrity of all constructs in *Agrobacterium* was confirmed by restriction analyses.

Plant Transformations

Transgenic tobacco plants cv. Xanthi were obtained through leaf disc transformation (Horsch et al., 1985) with selection for kanamycin resistance at 100 mg/l. Transformation of the plants was further verified by observing segregation for kanamycin resistance in their progeny (Derol and Gardner, 1988). Plants appearing to have more than three unlinked loci segregating for kanamycin resistance were not included in the analyses.

GUS Assays

Fluorometric GUS assays were carried out in protein extracts obtained from mature dry seeds essentially as describe by Jefferson (1987). Each value shown in Figure 3 represents the average between two independent protein extracts (two replicates) for each individual plant. Fluorometric readings of seed extracts obtained from untransformed control plants either regenerated *in vitro* or obtained directly from seeds, were similar to the readings obtained with extraction buffer alone, and therefore were negligible. Protein concentration in the seed extracts was determined by the Bradford method (Bradford, 1976) using a kit purchased from BioRad and bovine serum albumin as standard. Data on GUS activity was statistically analyzed with the SAS-GLM software (SAS Institute Inc., 1987). Histochemical staining for GUS activity was carried out on sections of dry seeds cut with a scalpel blade or on intact dissected embryos stained for 3 hours, without any fixation step, exactly as described (Jefferson, 1987).

Results

S/e Promoters Contain Several Putative Regulatory Elements

To begin identifying putative regulatory motifs in the *S/e* promoters we sequenced portions of the 5'-flanking region from four *S/e* loci. As expected from our previous work (Calvo et al., submitted), the clones were grouped into two classes, based on DNA sequence identity in their 5'-flanking region (Figure 1). One class was formed by *S/e1* and *S/e4* and the other by *S/e2* and *S/e3* promoters. Since the sequence identity observed across the two classes was very low (25 %) the sequence alignments are shown separately. *S/e1* and *S/e4* promoters (Figure 1A) have 68 % overall sequence identity. Several putative regulatory sequences occur in the proximal (in the vicinity of the TATA box) promoter region of both the *S/e1* and *S/e4* clones. These sequences are underlined in Figure 1A and include the TATA box at +1 bp, two SphI boxes (Hattori et al., 1992) or RY repeats (Dickinson et al., 1988) with the sequence CATGCATG(A) at -19 and -106 bp (positions on the *S/e1* promoter in relation to the putative TATA box), and an ACGT motif (Foster et al., 1994) at - 89 bp. The SphI box at -106 bp in *S/e1* promoter is flanked by two inverted repeats with the sequence CCACAT. A similar situation occurs in the *S/e4* promoter with the exception that the CCACAT repeat upstream to the SphI box has the second A replaced by a G (CCACGT). On the other hand, the *S/e2* and *S/e3* proximal promoter regions (Figure 1B) have a variant (CAATGCATG) of the SphI box and appear to be devoid of the CCACAT repeats. Moreover, the hexameric G-box core (CACGTG) is present in both the *S/e2* and *S/e3* proximal promoter region.

We extended our sequence analysis further upstream in the *S/e1* promoter (Figure 1A). Potentially significant motifs are gathered between positions -1250 to -1106 bp. These include two 15 bp tandem repeats containing the G-box core motif (at -1218 and 1250 bp), a SphI box variant (CATGACATG) at -1141 bp, and an ACGT core motif at -1110 bp. Sequence of another 496 bp upstream uncovered two more potential ACGT core motifs at positions -1541 and -1714 bp. Therefore, the distal (further upstream to the TATA box) region of the *S/e1* promoter contains the G-box core and SphI box-like motifs found in the proximal *S/e2* and *S/e3* promoters.

The results on the sequence analysis of the *S/e1-4* promoters are summarized in Figure 2, where we present a schematic diagram of the promoter region from *S/e1-4* along with the corresponding promoter region of all known *S/e* homologs from other plant species. Also shown are the promoters from two major storage protein genes (*Gy2* and *Gma α '*) and a Group3 Lea gene (*GmMP9*) from soybean. Unlike most of its homologs, the *S/e* promoters have two SphI boxes (*S/e1* and *S/e4*) or SphI box-like sequences (*S/e2* and *S/e3*) in the proximal promoter region which are characteristic of the promoter of several seed-specific genes including the soybean storage proteins (Dickinson et al., 1988). Furthermore, the *S/e1* promoter also distinguishes from the promoter from all known homologs by the absence of the G-box core in the proximal promoter region.

The Role of the Distal G-Box Core Motifs

We investigated the role of several putative regulatory motifs described above for the *S/e1* promoter. Transcription fusion of the *S/e1* promoter with the

GUS reporter gene were developed and introduced into tobacco plants. Rather than study unidirectional deletion derivatives of the *S/e1* promoter, we took the approach of analyzing the effect of specific mutations in the context of a 1823 bp promoter fragment. Our first question was to whether the sequences GTCACGTGTCATGGT present as two 15 bp direct repeats at positions -1218 and -1250 bp (with the first T of the putative TATA box as position +1; Figure 1A) play a key role in regulating *S/e1* expression. Our interest in these sequences derived from the fact that they contain the hexameric palindromic CACGTG motif or G-box core found in the promoter of several *Lea* genes as part of the ABRE sequence motifs (Quatrano et al., 1993). Two plasmid constructs were made to investigate the role of these sequences. The first, p Δ HincIIGUS, contains a 236 bp deletion and encompasses both repeats (Figure 3). Results shown in Figure 3 indicate that the amount of GUS activity accumulated in mature dry seeds of tobacco plants harboring this mutated promoter did not differ significantly from that observed for the control plants carrying the intact *S/e1* promoter fragment (p2618GUS). Since the p Δ HincIIGUS construct carries a rather large deletion, one could speculate that a possible positive or negative effect of the CACGTG motif on gene expression could have been counteracted by the deletion of other unknown sequence motif(s) elsewhere in the 236 bp fragment. Therefore, we created a second plasmid construct, p Δ PmlIGUS, which has a much smaller deletion (34 bp) that also eliminates both CACGTG core sequences (Figure 3). The results shown in Figure 3 indicate that plants carrying plasmid p Δ PmlIGUS also did not have GUS activity levels significantly different from the p2618GUS control plants. Taken together with the results from construct p Δ HincIIGUS, these results indicate that the CACGTG motifs at position -1216 and -1248 bp do not play a major role in regulating *S/e1* promoter activity in transgenic tobacco.

The Role of the Proximal ACGT Core Motif

Besides the CACGTG motifs, other ACGT-containing sequences are present in the *S/e1* promoter (Figure 1A; positions -89; -1110; -1541 and -1714 bp) .

Rather than test all four motifs, we selected the ACGT core at position -89 bp for several reasons. First, it is the one closest to the TATA box (Figure 1A). Second, it is found in the vicinity of other putative regulatory sequences (Figure 1A), and third, it is also present at a position in the promoter similar to other seed-expressed genes, including *Emp-1*, the *S/e* homolog in rice (Litts et al., 1992) (see Figure 5A and 5B). Therefore we reasoned that the ACGT core at -89 bp would be a more likely candidate to play a significant role in regulating *S/e1* promoter activity. We tested the role of the ACGT core at -89 bp by creating plasmid pMP3GUS in which the ACGT sequence was mutated to CAGG (Figure 3) through PCR-assisted mutagenesis (Landt et al., 1990). Nucleotide replacements in the ACGT core have been previously shown to eliminate efficient *in vitro* binding of bZIP proteins to target DNA sites containing ACGT core motif (Guiltinan et al., 1990; Armstrong et al., 1992; Schmidt et al., 1992; Schindler et al., 1992; Izawa et al., 1993) and also to prevent expression *in vivo* (Ueda et al., 1992). The results presented in Figure 4 show that, on average, plants harboring the mutated version of the promoter (pMP3GUS) have GUS activity levels in the dry seeds similar to the control (p2618GUS) plants. Thus, we conclude that it is unlikely that the ACGT core at position -89 bp is a major site for *S/e1* promoter regulation in transgenic tobacco plants.

The Role of the SphI Box Motif

Next, we investigated whether the SphI box present at position -106 bp in the *S/e1* promoter is important for promoter activity, as has been reported for other seed-specific genes including the two major seed storage proteins expressed during the maturation phase of soybean seed development (Chamberland et al., 1992; Lelievre et al., 1992). For this purpose, we developed plasmid pΔSphIGUS in which four nucleotides of the CATGCATG repeat were deleted (Figure 3). Unlike the previous mutations of the *S/e1* promoter described above, disruption of the SphI box significantly reduced GUS activity in mature seeds when compared to control plants harboring the p2618GUS plasmid (Figure 4). This result supports a positive role for the SphI box in *S/e1* promoter activity in tobacco seeds.

The Role of the CCACAT Repeats

The *S/e1* promoter contains two 6 bp motifs (CCACAT) flanking the SphI box as inverted repeats (at positions -96 and -115 bp; Figure 1A). This observation prompted us to test whether the CCACAT repeats are also important for proper *S/e1* promoter function. To do this we created plasmid pMP1-MP2GUS (Figure 3) in which 7 bp nucleotide replacements were introduced in both inverted repeats through PCR-assisted mutagenesis. The results shown in Figure 4 clearly show that mutations in the CCACAT repeats had a dramatic negative quantitative effect on the accumulation of GUS activity in tobacco mature seeds when compared to the p2618GUS control plants.

Due to the importance of the CCACAT repeats for *S/e1* promoter function, we conducted a database search for this sequence in the vicinity of the SphI boxes

of other seed-expressed genes. The results of this search are presented in Table II where we list all the thirty-one genes searched and the position of their respective SphI boxes. We initially conducted this search by allowing 1 mismatch anywhere in the CCACAT sequence. Twenty-nine genes fulfilled this criteria. However, only five genes, in addition to the *S/e1* and *S/e4*, were found to contain an identical CCACAT sequence at 60 bp or less from the SphI box and they were designated class 1 in Figure 5A. Two of these five are also Lea genes, the rice *Emp-1* (Litts et al., 1992) and the maize *Rab-17* (Vilardell et al., 1990). Three other non-Lea genes were also found to contain the CCACAT sequence; the maize zein *pML1* and the *Vicia faba* USP and *LeB1* genes. Thus, the apparent conservation of this sequence across species as distant as rice and soybean supports the idea that this motif may play a significant role in gene regulation during seed development. Interestingly, *pML1* is a pseudogene (Wandelt and Feix, 1989) and the CCACAT sequence in its promoter is a variant of the CCACGT sequence which is part of the opaque-2 binding site in the maize zein promoter (Figure 5, compare *pML1* with 22Z-4 sequences). Moreover, the *LeB1* from *Vicia faba* is also thought to be a pseudogene (Heim et al., 1989). However, unlike the other members of the legumin gene family, the *LeB1* promoter has a 16 bp deletion which includes the SphI box motif.

The other twenty-four genes have a CCACAT-like (1 mismatch allowed) motif and were designated class 2 genes. Some examples are shown in Figure 5B. An alignment of the CCACAT-like motif of the class 2 genes (date not shown) revealed that twenty-two of them share a common CCAC sequence.

The two remaining genes which appear to be devoid of the CCACAT-like sequence nearby the SphI box were designated class 3 genes. Examples of these genes are shown in Figure 5C.

Histochemical Staining of Seeds

Finally, we examined the spatial pattern of GUS activity conferred by the *S/e1* promoter in sectioned dry seeds of transgenic tobacco plants. Results of a typical staining pattern are shown in Figure 5. GUS staining was easily detected in seeds from the p2618GUS control plants in both the embryo and the endosperm. Generally, the whole embryo was intensely stained and the staining in the endosperm was less intense.

Discussion

The apparent ABA-independent regulation of the *S/e1* gene from soybean (Calvo et al., submitted), together with the fact that promoter analyses of legume Lea genes has not been previously reported, led us to analyse the *S/e1* promoter region in a search for cis-acting elements involved in its regulation. The results presented here show that, unlike the CACGTG core at -1216 and -1248 bp and the ACGT core at -89 bp, the SphI box at position -106 bp and its flanking CCACAT repeats have a major positive quantitative effect on *S/e1* promoter function in transgenic tobacco seeds. To our knowledge, this is the first report of the involvement of these sequences in plant Lea gene expression.

Of the ACGT-containing sequences in the *S/e1* promoter, the CACGTG at positions -1216 and -1248 bp are the sequences that most resemble the ABRE motif commonly found in the promoters of Lea and storage protein genes. ABRE have been shown to confer ABA-response and proper quantitative expression throughout seed development (Marcotte et al., 1989; Guiltinan et al., 1990; Thomas, 1993; Kawagoe et al., 1994). Nonetheless, we provide evidence that the

CACGTG sequences in the *S/e1* promoter are not essential for correct quantitative expression in transgenic tobacco. Similarly, the conserved (see Figure 5A and 5C) proximal ACGT core located at -89 bp does not seem to be required for quantitative expression of *S/e1* promoter in tobacco seeds. At least two explanations could account for this discrepancy. First, the CACGTG sequences in the *S/e* promoter are located considerably distant from the TATA box in comparison to the previously characterized ABREs (Guiltinan et al., 1990; Pla et al., 1993), and thus they might be expected to have a lesser impact in promoter activity. Second, the nucleotides flanking the CACGTG core could account for this difference since they have been shown to affect the specificity of protein binding *in vitro* (Williams et al., 1992; Izawa et al., 1993) and the pattern of expression *in vivo* (Salinas et al., 1992). Our results are in agreement with the previous observation that *S/e1* promoter seem to be less ABA-dependent than the promoter of *S/e* homologs from other species (Calvo et al., submitted) and they emphasize the risk of attributing a major regulatory function to a G-box core solely on the basis of sequence identity (Williams et al., 1992). Nevertheless, as is the case with promoter analyses performed in a heterologous host, it is possible that the G-box core and the ACGT core at -89 bp are major sites for *S/e1* regulation in soybean plants but could not function similarly in tobacco.

Since both positive (Bäumlein et al., 1992; Chamberland et al., 1992; Lelievre et al., 1992) and negative effects (Fiedler et al., 1993) on seed-specific gene expression have been observed for the SphI box or RY repeat, we could not previously predict a possible role for this motif in the *S/e1* promoter. Previous work has suggested a requirement of the SphI box for proper ABA regulation and VP1 transactivation of the *C1* promoter in maize aleurone (Hattori et al., 1992). Our results extend the importance of the SphI box to an ABA-independent gene expressed in the embryo of transgenic tobacco seeds.

We can not conclusively rule out that the positive effect conferred by the SphI box was caused by a change in the spatial relationship of other surrounding cis-acting elements resulting from the 4 bp deletion in p Δ SphIGUS. However, both nucleotide replacements (Chamberland et al., 1992) and deletions (Lelievre et al., 1992) resulted in the same qualitative effect in GUS expression driven by two different soybean storage protein promoters. A similar result was also observed for a 4 and a 2 bp deletion in the SphI box of the USP promoter of *Vicia faba* (Fiedler et al., 1993).

The observation that the SphI box has the same positive role in gene expression in the promoter of genes expressed during either the mid (β -conglycinin and glycinin promoters; Chamberland et al., 1992 and Lelievre et al., 1992, respectively) or the late phases (*S/e1*; this paper) of soybean seed development raises the question of whether or not this effect is accomplished by interaction of the SphI box with the same transcription factor(s) during both phases of seed development. If so, how would they act in order to bring about two very distinct patterns of gene expression in the same cells? One possible scenario would be that the temporally differential activation of both target genes would be accomplished through interactions mediated by other yet unknown sequence motif(s) that, unlike the SphI box, are not present in the promoter of both genes. Indeed, preliminary evidence suggests that the positive effect of the SphI box in the promoter of the *Vicia faba* legumin gene depends on other sequences (Bäumlein et al., 1992). Further, the VP1-mediated ABA response of the maize *C1* promoter depends on both the SphI box and an adjacent sequence (Hattori et al., 1992).

Our initial hypothesis was that the CCACAT inverted repeats could be a new sequence motif involved in the distinction of these two temporal programs in soybean embryos. Although we have measured GUS activity only in mature

seeds, it has been previously observed that even when driven by storage protein promoters, the GUS activity measured in dry seeds is an accurate reflection of activity throughout the "maturation" phase of embryogenesis (Chamberland et al., 1992). So, it is unlikely that the mutations on the CCACAT repeats are simply bringing GUS transcription to an earlier phase (i.e. the "maturation" phase) of seed development since GUS activity was drastically reduced. Therefore, we speculate that if the CCACAT repeats are involved in the coupling of the SphI enhancing effect to late embryogenesis it does so in a way that loss of its function does not restore gene expression to earlier stages of seed development. Since not all soybean *Lea* genes seem to contain SphI boxes and CCACAT motifs, other mechanism(s) capable of conferring expression at late embryogenesis must exist in soybean embryos also. Moreover, according to this view it would be difficult to explain the existence of the CCACAT motif in non-*Lea* genes (Figure 5C). It is quite intriguing though, that from the three non-*Lea* genes where the CCACAT sequence was found, two are pseudogenes (*pML1* and *LeB1*) and in the third one (*USP*) the SphI box was found to have an unusual negative effect in gene expression. The complete analyses of the CCACAT mutant plants along with gain of function experiments and analyses of other legume *Lea* genes should shed more light onto these question.

A second possibility is that, rather than representing a new sequence motif, the CCACAT repeats, in the context that they usually occur (see Figure 5C), are only a variant of the CACGTG core or of the more recently described CANNTG (E-box; Kawagoe and Murai, 1992) motifs competing for the same transcription factors. For instance, a simple replacement of the second adenine for a guanine restores a CACGTG core in the *S/e1* promoter. Indeed, this variant does occur in the promoter of the *S/e4* pseudogene (Figure 5C). Evidence against this

hypothesis come from studies with the transactivation of the different zein promoters by the maize opaque-2 bZIP. Replacement of an ACGT by an ACAT core, a variant that also occurs naturally (see Figure 5C), severely reduces both *in vitro* protein binding (Schmidt et al., 1992) and transactivation *in vivo* (Ueda et al., 1992). Proteins that are able to bind the E-box, and therefore unlike most cloned bZIP proteins (Foster et al., 1994), are unable to distinguish the internal CG dinucleotides of the CACGTG motif, have been described in seed nuclear extracts (Kawagoe and Murai, 1992). However, their presence during late embryo development awaits further investigation. Moreover, this hypothesis would not explain the conservation of the CCACAT repeats located downstream to the SphI box on the *S/e1* and *S/e4* promoters. Protein binding experiments on these promoter as well as the analyses of the effect of these mutations *in vivo* should allow us to answer this question more conclusively.

While this manuscript was in preparation, Shen and Ho (1995) reported a novel sequence motif, CE1, necessary along with a G-box element, to confer ABA response in barley aleurone and which may also be involved in the ABA response of several *Lea* genes. Sequence conservation among the putative CE1-like motifs is restricted to the central CACC core which are contained in the CCACAT repeats described in this study. It will be interesting to determine if these motifs are at all related and whether they may be part of a new family of motifs having a common CCAC core sequence.

The *S/e1* promoter is able to drive GUS gene expression in both the embryo and, at lower levels, in the endosperm of tobacco seeds. These results contrast with those obtained when the promoter of the *S/e* homolog in wheat, *Em*, was used to drive the GUS gene expression in tobacco (Marcotte et al., 1989). These authors were able to observe GUS activity only in the embryo of tobacco seeds.

Also, expression of *S/e* homologs in both maize (Williams and Tsang, 1991) and carrot (Wurtele et al., 1993) appears to be confined to the embryo. We presently do not know the cause of this discrepancy. Nevertheless, unlike all the monocots and several dicots, including carrot, the soybean endosperm is very short lived and late embryo development occurs in the absence of endosperm (Chamberlin et al., 1994). Thus, the lack of a selection pressure for precise spatial distinction between endosperm and embryo expression may account for the retention of the endosperm expression by the *S/e1* promoter. The analyses of the spatial pattern of expression in the seeds of tobacco plants carrying the mutated versions of the *S/e1* promoter may provide some insight into the regulatory sequences conferring endosperm expression.

References

- Armstrong, G.A., Weisshaar, B., Hahlbrock, K. (1992) Homodimeric and heterodimeric leucine zipper proteins and nuclear factors from parsley recognize diverse promoter elements with ACGT cores. *Plant Cell* 4: 525-537.
- Baker, J., Steele, C., III L.D. (1988) Sequence and characterization of 6 *Lea* proteins and their genes from cotton. *Plant Mol. Biol.* 11: 277-291.
- Bäumlein, H., Nagy I., Villarroel, R., Inze D., Wobus, U. (1992) Cis-analysis of a seed protein gene promoter the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2: 233-239.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.
- Calvo, E.S., Roderme, I. S., Shoemaker, R.S. (1994) A third highly conserved group 1 *Lea* gene from *Arabidopsis thaliana*. *Plant Physiol.* 106: 787-788
- Chamberland, S., Daigle, N., Bernier, R. (1992) The legumin boxes and the 3' part of a soybean β -conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. *Plant Mol. Biol.* 19: 937-949.

- Cho, T.J., and Nielsen, N.C. (1989) The glycinin Gy3 gene from soybean. *Nucl. Acids. Res.* 17: 4388-4388.
- Conceição, A.S., Krebbers, E. (1994) A cotyledon regulatory region is responsible for the different spatial expression patterns of *Arabidopsis* 2S albumin genes. *Plant J.* 5: 493-505.
- Derol, S., Gardner, R. (1988) Expression and inheritance of kanamycin resistance in a large number of transgenic petunias generated by *Agrobacterium* mediated transformation. *Plant Mol. Biol.* 11: 355-364.
- Dickinson, C.D., Evans, R.P., Nielsen, N.C. (1988) RY repeats are conserved in the 5' flanking region of legume seed protein genes. *Nucl. Acids. Res.* 16: 371.
- Doyle, J.J., Schuler, M.A., Godette, W.D., Zenger, V., Beachy, R.N., Slightom, J.L. (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. *J. Biol. Chem.* 261: 9228-9238.
- Fiedler, U., Filistein, R., Wobus, U., Baumlein, H. (1993) A complex ensemble of *cis*-regulatory elements controls the expression of a *Vicia faba* non-storage seed protein gene. *Plant Mol. Biol.* 22: 669-679.
- Finkelstein, R.R. (1993) Absciscic acid-insensitive mutations provide evidence for stage-specific signal pathways regulating expression of an *Arabidopsis* late embryogenesis-abundant (*lea*) gene. *Mol. Gen. Genet* 238: 401-408.
- Foster, R., Izawa T., Chua, N.-H. (1994) Plant bZIP proteins gather at ACGT elements. *FSEB J.* 8: 192-200.
- Fujiwara, T., Beachy, R.N. (1994) Tissue-specific and temporal regulation of a β -conglycinin gene: roles of the RY repeat and other *cis*-acting elements. *Plant Mol. Biol.* 24: 261-272.
- Fukazawa, C., Momma, T., Higuchi, W., Udaka, K. (1987) Complete nucleotide sequence of the gene encoding a glycinin A(2)B(1a) subunit precursor of soybean. *Nucl. Acids Res.* 15: 8117-8117.
- Galau, G.A., Jakobsen, K.S., Hughes, D.W. (1991) The controls of late dicot embryogenesis and early germination. *Phys. Plant.* 81: 280-288.
- Galau, G.A., Wang, H.Y-C., Hughes, D.W. (1992) Cotton *Lea4* (D19) and *LeaA2* (D132) group 1 *Lea* genes encoding water stress-related proteins containing a 20-amino acid motif. *Plant Physiol.* 99: 783-788 (1992).
- Gallusci, P., Salamini, F., Thompson, R.D. (1994). Differences in cell-type-specific expression of the gene *Opaque-2* in maize and transgenic tobacco. *Mol. Gen. Genet.* 244: 391-400.

- Gaubier, P., Raynal, M., Hull, G., Huestis, G.M., Grellet, F., Arenas, C., Pages, M., Delseny M. (1993) Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Mol Gen Genet.* 238: 409-418.
- Guiltinan, M.J., Marcotte, Jr.W.R., Quatrano, R.S. (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* 250: 267-271.
- Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C., McCarty, D.R., Vasil, I.K. (1992) The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev.* 6: 609-618.
- Heim, U., Schubert, R., Bämlein, H., Wobus, U. (1989) The legumin gene family: structure and evolutionary implications of *Vicia faba* B-type and pseudogenes. *Plant Mol. Biol.* 13: 653-663.
- Hetherington, A.M. and Quatrano, R.S. (1991) Mechanisms of action of abscisic acid at the cellular level. *New Phytol.* 119: 9-32.
- Hoffman, L.M. and Donaldson D.D. Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* 4: 883-889.
- Hollung, K., Espelund, M., Jakobsen, K.S. (1994) Another *Lea* B19 gene (Group1 *Lea*) from barley containing a single 20 amino acid hydrophilic motif. *Plant Mol. Biol.* 25: 559-564.
- Horsch, R.B., Fry J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G., Fraley, R.T. (1985) A simple and general method for transferring genes into plants. *Science* 227: 1229-1231.
- Izawa, T., Foster, R., Chua, N.-H. (1993) Plant bZIP protein DNA Binding Specificity. *J. Mol. Biol.* 230: 1131-1144.
- Jefferson, R.A. (1987) Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Molec. Biol. Rep.* 5: 387-405.
- Josefsson, K.D., Lenman, M., Ericson, M.L., Rask, L. (1987). Structure of a gene encoding the 1.7S storage protein, napin, from *Brassica napus*. *J. Biol. Chem.* 262: 12196-12201.
- Kawagoe, Y., Campell, B.R., Murai, N. (1994) Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein β -phaseolin gene. *Plant J.* 5: 885-890.

- Kawagoe, Y. and Murai, M. (1992) Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein β -*phaseolin* gene conferring spatial and temporal control. *Plant J.* 2: 927-936.
- Landt, O., Grunert, H.-P., Hahn, U. (1990) A general method for rapid site-directed mutagenesis using the polymerase chain reaction. *Gene* 96: 125-128.
- Lelievre, J.-M., Oliveira, L.O., Nielsen, N.C. (1992) 5'-CATGCAT-3' elements modulate the expression of glycinin genes. *Plant Physiol.* 98: 387-391.
- Lee, P., Chow, T., Chen, Z., Hsing, Y.C. (1992). Genomic nucleotide sequence of a soybean seed maturation protein GmPM9 gene. *Plant Physiol.* 100: 2121-2122.
- Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, D.M., Boulter, D. (1986). The 5'-flanking regions of three pea legumin genes; comparison of the DNA sequences. *Nucl. Acids Res.* 13: 6733-6743.
- Litts, J.C., Colwell, G.W., Chakerian, R.L., Quatrano, R.S. (1991) Sequence analysis of a unctonal member of the *Em* gene family from wheat. *DNA Sequence-J. DNA Sequencing and Mapping* 1: 263-274.
- Litts, J.C., Erdman, M.B., Huang, N., Karrer, E.E., Noueiry, A., Quatrano, R.S., Rodriguez, R.L. (1992) Nucleotide sequence of the rice (*Oryza sativa*) *Em* protein gene (*Emp1*). *Plant Mol. Biol.* 19: 335-337.
- Liu, C.-N. and Rubenstein, I. (1992). Genomic organization of an α -zein gene cluster in maize. *Mol. Gen. Genet.* 231: 743-751.
- Marcotte, Jr.W.R., Russell, S.H., Quatrano, R.S. (1989) Absciscic acid-responsive sequences from the *Em* gene of wheat. *Plant Cell* 1: 969-976.
- Mundy, J., Yamaguchi-Shinozaki, K., Chua, N.-C. (1990) Nuclear proteins bind conserved elements in the absciscic acid-responsive promoter of a rice *rab* gene. *Proc. Natl. Acad. Sci. USA.* 87: 1406-1410.
- Nielsen, N.C., Dickinson, C.D., Cho, T., Thanh, V.H., Scallan, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., Goldberg, R.B. (1989) Characterization of the glycinin gene family in soybean. *Plant Cell* 1: 313-328.
- Ottoboni, L.M.M., Leite, A., Targon, M. L. P. N., Silva, M. J., Arruda, P. (1993) Sequence analysis of 22KDa-like α -coixin genes and their comparison with homologous zein and kafirin genes reveals highly conserved protein structure and regulatory elements. *Plant Mol. Biol.* 21: 765-778.

- Paz-Arez, J., Ghosal, D., Wienand, U., Peterson, A. P., Saedler, H. (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* 6:3553-3558.
- Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., Giraudat, J. (1994) Regulation of gene expression programs during *Arabidopsis* seed development: role of the *ABI3* locus and of endogenous abscisic acid. *Plant Cell* 6: 1567-1582.
- Pla M., Vilardell, J., Guiltinan, M.J., Marcotte, Jr. W.R., Niogret, M.F., Quatrano, R.S., Pages M. (1993) The *cis*-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene *rab28*. *Plant Mol. Biol.* 21: 259-266.
- Quatrano, R.S. (1986) Regulation of gene expression by abscisic acid during angiosperm embryo development. *Oxford Surveys Plant. Mol. Cell Biol.* 3: 467-477.
- Quatrano, R.S., Marcotte, Jr.W.R., Guiltinan, M. (1993) Regulation of gene expression by abscisic acid In: Ed. Verma *DPS Control of Plant Gene Expression*, 69-90. CRC Press, Boca Raton, FL.
- SAS/STAT guide for personal computers. Version 6 edition, Cary, NC.
- Salinas, J., Oeda, K., Chua, N.-H. (1992) Two G-box-related sequences confer different expression patterns in transgenic tobacco. *Plant Cell* 4: 1485-1493.
- Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- Scallan, B.J., Dickinson, C. D., Nielsen, N.C. (1987). Characterization of a null-allele for the *Gy4* glycinin gene from soybean. *Mol Gen. Genet.* 208: 107-113.
- Schindler, U., Terzaghi, W., Beckmann, H., Kadesch, T., Cashmore, A.R. (1992) DNA binding site preferences and transcriptional activation properties of the *Arabidopsis* transcription factor GBF1. *EMBO J.* 11: 1275-1289.
- Schmidt, R.J., Burr, F.A., Aukerman, M.J., Burr, B. (1990) Maize regulatory gene *Opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc. Natl. Acad. Sci. USA.* 87: 46-50.
- Schmidt, R.J., Ketudat, M., Aukerman, M.J., Hoschek, G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-KD zein genes. *Plant Cell.* 4: 689-700.

- Shen, Q. and Ho, T.-H.D. (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. *Plant Cell* 7: 295-307.
- Sims, T.L. and Goldberg R.B. (1989) The glycinin *Gy1* gene from soybean. *Nucl. Acids Res.* 17: 4386-4386.
- Skriver, K. and Mundy, J. (1990) Gene expression in response to abscisic acid and osmotic stress. *Plant Cell* 2: 503-512.
- Thanh, V.H., Tumer, N.E., Nielsen, N.C. (1989) The glycinin *Gy2* gene from soybean. *Nucl. Acids Res.* 17: 4387-4387.
- Thomas, T.L. (1993) Gene expression during plant embryogenesis and germination. *Plant Cell* 5: 1401-1410.
- Thompson, J.A., Bown, D., Yaish, S., Gatehouse, J.A. (1991) Differential expression of seed storage protein genes in the pea *legJ* subfamily; sequence of gene *legK*. *Biochem. Physiol. Pflanzen* 187: 1-12.
- Ueda, T., Wawerczak, W., Ward, K., Sher, N., Ketudat, M., Schmidt, R., Messing, J. (1992) Mutations of the 22- and 27-KD zein promoters affect transactivation by the Opaque-2 protein. *Plant Cell* 4: 701-709.
- Vilardell, J., Goday, A., Freire, M.A., Torrent, M., Martinez, M.C., Torne, J.M., Pages, M. (1990) Gene sequence, developmental expression, and protein phosphorylation of RAB-17 in maize. *Plant Mol. Biol.* 14: 423-432.
- Vilardell, J., Martínez-Zapater, J.M., Goday, A., Arenas, C., Pagés, M. (1994) Regulation of the *rab17* gene promoter in transgenic *Arabidopsis* wild-type, ABA-deficient and ABA-insensitive mutants. *Plant Mol. Biol.* 24: 561-569.
- Wandelt, C. and Feix, G. (1989) Sequence of a 21-Kd gene from maize containing an in-frame stop codon. *Nucl. Acids Res.* 17: 2354.
- Williams, M.E., Foster, R., Chua, N.-H. (1992) Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. *Plant Cell* 4: 485-496.
- Wurtele, E.S., Wang, H., Durgerian, S., Nikolau, B.J., Ulrich, T. (1993) Characterization of a gene expressed early in somatic embryogenesis of *Daucus carota*. *Plant Physiol.* 102: 303-312.

Table I. Synthetic oligonucleotides used in PCR-mutagenesis.

Name	Sequence ¹	Position ²
MP1	5'-CTTTCGCCTAAAGCCtCgaGTCCATGCATGCC-3'	-128 to -97
MP2	5'-GTCCATGCATGCCTctAgaTACGTTTCCATCAAAATC-3'	-109 to -72
MP3	5'-GCATGCCTACACCTcaGgTTCCATCAAAATCTAGTC-3'	-103 to -68

¹ Bases replaced are shown in lowercase letter.

² Numbers correspond to *S/e1* promoter and are relative to the putative TATA box.

Table II. Seed-expressed genes containing the SphI box.

Gene	Organism	Position ¹	Class ²	Reference ³
<i>pML1</i>	<i>Z. mays</i>	-218	1	Wandelt & Feix, 1989
<i>Rab-17</i>	"	-75	1	Villardell et al., 1990
<i>LeB1</i> [*]	<i>P. vulgaris</i>	-67	1	Heim et al., 1990
<i>USP</i>	<i>V. faba</i>	-145	1	Fiedler et al., 1993
<i>Emp-1</i>	<i>O. sativa</i>	-323	1	Litts et al., 1992
<i>LegJ</i>	<i>P. sativum</i>	-67	2	Thompson et al., 1991
<i>LegK</i>	"	-67	2	"
<i>LegA</i>	"	-77	2	Lycett et al., 1986
<i>AT2S1</i>	<i>A. thaliana</i>	-47	2	Conceição et al., 1994
<i>AT2S2</i>	"	-43	2	"
<i>AT2S3</i>	"	-48	2	"
<i>napA</i>	<i>B. napus</i>	-44	2	Josefsson et al., 1987
<i>LeB4</i>	<i>V. faba</i>	-67	2	Bäumlein et al., 1992
<i>Pvuβ</i>	<i>P. vulgaris</i>	-38; -228	2	Doyle et al., 1986
<i>DLEC1</i>	"	-80; -140	2	Hoffman & Donaldson, 1986
<i>DLEC2</i>	"	-90; -185	2	"
<i>C1</i>	<i>Z. mays</i>	-115; -232	2	Paz-Aeres et al., 1987
<i>22Z4</i>	"	-208	2	Schmidt et al., 1990
<i>ZSF4C3</i>	"	-211	2	Liu & Rubenstein, 1992

Table 2 cont.

<i>α-3B</i>	Coix	-213	2	Ottoboni et al., 1993
<i>Em</i>	<i>T. aestivum</i>	-204; -359	2	Litts et al., 1991
<i>Gmaα¹</i>	<i>G. max</i>	-25; -212	2	Doyle et al., 1986
<i>Gy1</i>	"	-77; -221	2	Sims & Goldberg, 1989
<i>Gy2</i>	"	-76; -177	2	Thanh et al., 1989
<i>Gy3</i>	"	-77; 224	2	Cho & Nielsen, 1989
<i>A2B1a</i>	"	-76	2	Fukazawa et al., 1987
<i>sslwsrp</i>	"	-104	2	Z18359
<i>oleosinB</i>	"	-43	2	U09119
<i>GB-D-II</i>	"	-218	2	x68707
<i>O2</i>	<i>Z. mays</i>	-102; -159	3	Gallusci et al., 1994
<i>Gy4</i>	<i>G. max</i>	-122	3	Scallan et al., 1987

¹Positions in relation to the TATA box.

²Classes were established based on the presence or absence of a CCACAT or CCACAT-like motif nearby the SphI box.

³In case of unpublished genes the GenBank accession number is given.

* *LeB1* promoter has a 16 bp deletion that includes the SphI box motif.

A.

<i>Sle1</i>	AAGCTTTCCCGCCGGCAACAATGGCGGCATCC <u>ACGTCC</u> ATAACAAACCCC	-1697
	GTCTGTAATAAGTTTGAAAAGAAACCCCTCTTGGTAAAATGTTTCTAAAA	-1647
	CAAGCCCTCTGAAGTAAATTTGGCTACAACCAGTTCTCATAATATTATAA	-1597
	TTATTTTTTGAGAAAAAAATTATGCCCTGATACTATAAATAATATATAAT	-1547
	CATTA <u>ACGTCT</u> TATTTTAACCTGAAAACATGGTTTTCTCAACTTCTCATT	-1497
	ACAATCACTTTTAAGCATTAGTATCTATTCCTTAACTTTTAATAGAAGG	-1447
	CGAAAAGTAATATCTTTGTCAACCATGGTTAATTTTAGACAAAATTATAT	-1397
	TTTTGGTCCTCCAGTTTATCTCCAGTTTGGATTAGTTTCTTATTAATT	-1347
	TAATTCACCAATTGAGTCTTCCTATTTTGTAATAATCGTGCAATGTTGATC	-1297
	TCTCAGACCACAATTAGACATTGATCGTTAGCAAGTGACATTGACGGTCA	-1247
	<u>CGTGTCATGTTTTT</u> ATTAGATGATGACTGTCACGTGTCATGTTCTGATTG	-1197
	GATGTCAACCTCATGAAAGATGAAATGCATTGTTGTTTTATTGACCACT	-1147
	CGAAACATGACATGTGACAGTTATCATCCAATAAGA <u>ACGTG</u> ACACATGGC	-1097
	AGGTAATATTTTTTTGGAACGGGCAATGTCCAATGTGGTCTAGGGAGCCA	-1047
	ACATTGCACGATTTTAAAAATACTCCTGGGTTAAGCACTTAATTTAATA	-997
	ACCCCTCTAGTTTAGGCTTTAATTTCTATTTCTAAGGGACAACAAATACAA	-947
	TTTTGGCTTAATTTTAACCATAGCATCATTAATATAATAACTGGACAAAT	-897
	TTAGTGTAACAAGTCATATGTGGCATCCTATGGCAATCTTCATTAGTCAC	-847
	ACTACTCACTACTCATTGACAGTTTTGGTGAATAGAGAGGAAATCTTTTC	-797
	TTGTCTTAGAAACGGGCCGAATCTAGTGAAGACAAGTTGAAGGTGTTGC	-747
	TAGGTATACAAGACATATTACACCCAACATAGAAAATAAATTTCTCAAAA	-697
	TGCCCTTTATAATAAACATTTATGGATATGGCATTTTATAATTGAAATTT	-647
	GTGTATGATATACTTAATTTTTTAAATACAAGTGAATTTTCAAGTTATAAAA	-597
	TGACATACCCATAAATGTTTATTTTATCCAAAATTAGATAGTCTTAATCA	-547
	TCCTTTAAAAAACAACCTAGGCTTGATCATCAAGAAAATATGACATAGT	-497
	TTAATAGTTAGATGTATTAAACTCAGATTTTTTTAAAGGAATTAACTCA	-447
	GAGTAATTGAGTGAATAAGTGACCAAGACTTACTCCTATCTATGTAAAT	-397
<i>Sle1</i>	CCCTTCCCAAAGAAGAGCAATATTTTAACAA-----TACATTTTAA--	-356
<i>Sle4</i>	GGTCGACCTGCAGGTCACG	-404

Figure 1. DNA sequence alignment of four *S/e* 5'-flanking sequences. Due to the low level of sequence identity between the two classes of genes (Calvo et al, submitted; see results section) the alignment is presented separately for the two classes. Numbers on the right designate the nucleotide position in relation to the putative TATA box of each gene. Sequence motifs discussed in the text are underlined. (A) Alignment between *S/e1* and *S/e4* promoters. (B) Alignment between *S/e2* and *S/e3* promoters.

S/e1	-----TCATAATATTAAT-----ATTATATGAAGNG--CGTCT--AGTCA	-319
S/e4	GATCTGTTTCCATAATCAATCCACCTAATTAATAATGTCGTTCTTARACCA	-354
S/e1	CGTATTTT-GGGCAGCTTGCNGCTATCATTCCTA-----AT-----TTAATTT	-279
S/e4	TACATTTCAAATTCAAATGGTAAATATATAATAATAGGGAATGCGATGTTATTA	-304
S/e1	GGATTCGTGGCGATTAAATACAGTAAGTAAGT-TCAACCTTATTCCTAA--A	-232
S/e4	GGTTCCTCAGTCACGAACGAATTGAGCAGCTTGGCTGTATATATATCATCA	-254
S/e1	TCACCTTAAATCTATCNGTTGTT-TTACAAAT-----TATATT	-196
S/e4	CAATTTATHTGCACCTTGTAGCGAATACAGTACACCTTGTGCATTTATATT	-204
S/e1	TC-GTATAACTCGTAAATGCTGCTC-----ANGATAT-----GCCA	-161
S/e4	CGTAAATCACATGAATATCAATCTGTTGGTTTACAAATATATTTAGTATAA	-154
S/e1	TTTTTCACT--EFGTCACACCTAAAGTGGGTGCTCTTTGGCCTTAAGCC	-114
S/e4	CTCGTATATATGCTGCTACTCTAAGTGGCTTGTGCTCCCTCCACCAATAGCC	-104
S/e1	ACATGTCCATGCAATGCCATACACCTACGTTTCCATCAAAATCTAGTCTTAT	-64
S/e4	AGGTGCCAATGCATGCCATACACCTACGTTTCCATCAAAATCTAGTCTTAT	-54
S/e1	CTTTTACTGGTAATCTTATTAATTAATTTGCAATTAGTTTACTACTTTACATGCA	+14
S/e4	CTTTTACTGGTAATCTTATTAATTAATTTGCAATTAGTTTACTACTTTACATGCA	+14
S/e1	TACTTTTACACACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	+34
S/e4	TACTTTTACACACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	+37
S/e1	TCTGACACATCGTGAAGTTTCTCTTTGAGAGAGTAAATGATCTTAGTGCA	+94
S/e4	TCTTACACATCGTGAAGTTTCTCTTTGAGAGAGTAAATGATCTTAGTGCA	+97
S/e1	ACTAATTAAC ATG +104	
S/e4	ACTAATTAAC ATG +107	

Figure 1 cont.

Figure 2. Schematic representation of the promoter region from Group 1 Lea genes. The data was compiled from published sequence from the following genes: *Arabidopsis AtEm 1* and *AtEm6* (Gaubier et al., 1993; Calvo et al., 1994), barley B19.1 (Hollung et al., 1994), carrot EMB-1 (Wurtele et al., 1993), cotton D19 and D 132 (Galau et al., 1992), rice *Emp-1* (Litts et al., 1992), wheat *Em* (Litts et al., 1991), and soybean *Slc1-4* (this study). The occurrence of similar motifs is also shown for the *Gy2* (Thanh et al., 1989) and *Gma α'* β -conglycinin (Doyle et al., 1986) storage protein and the GmPM9 (Lee et al., 1992) Group 3 Lea gene promoters from soybean. Arrows represent the direction (5' to 3') of the CCACAT sequence motif.

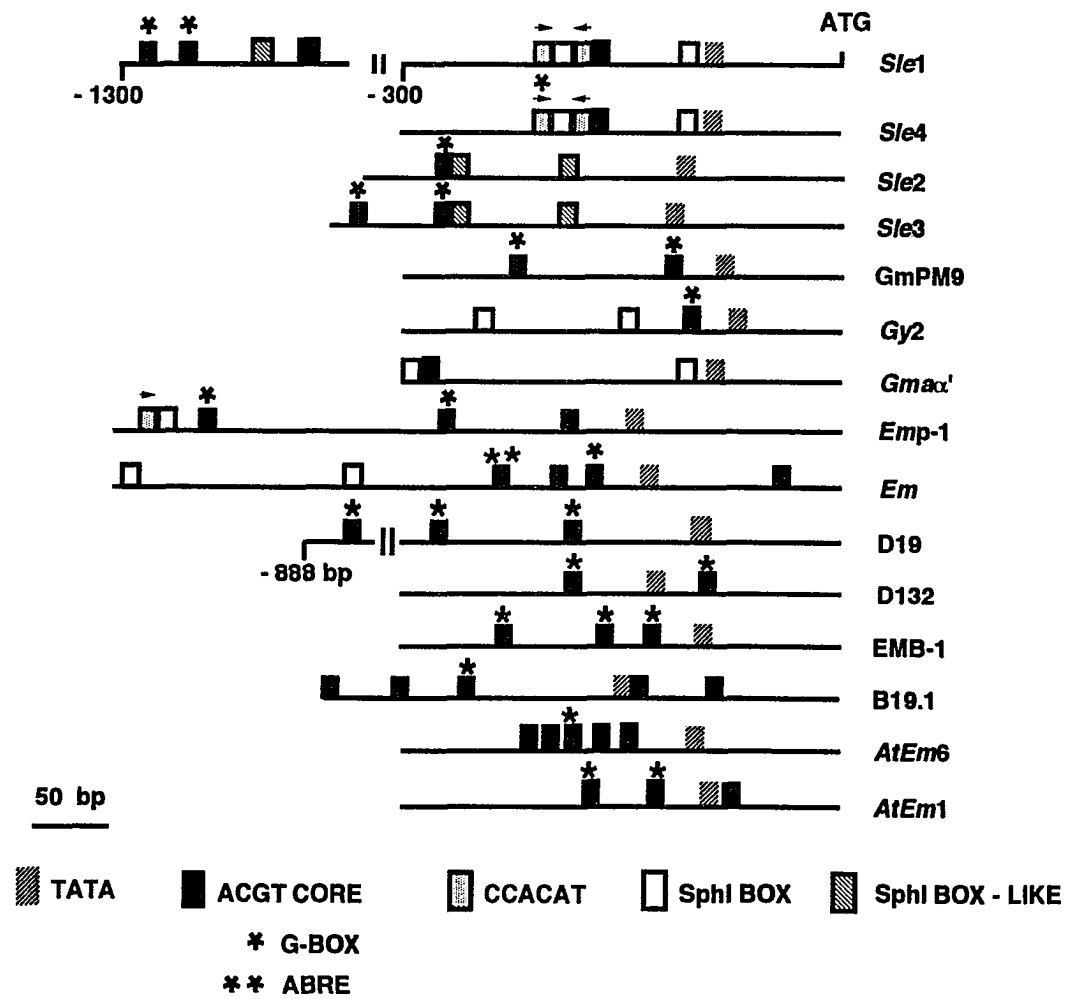
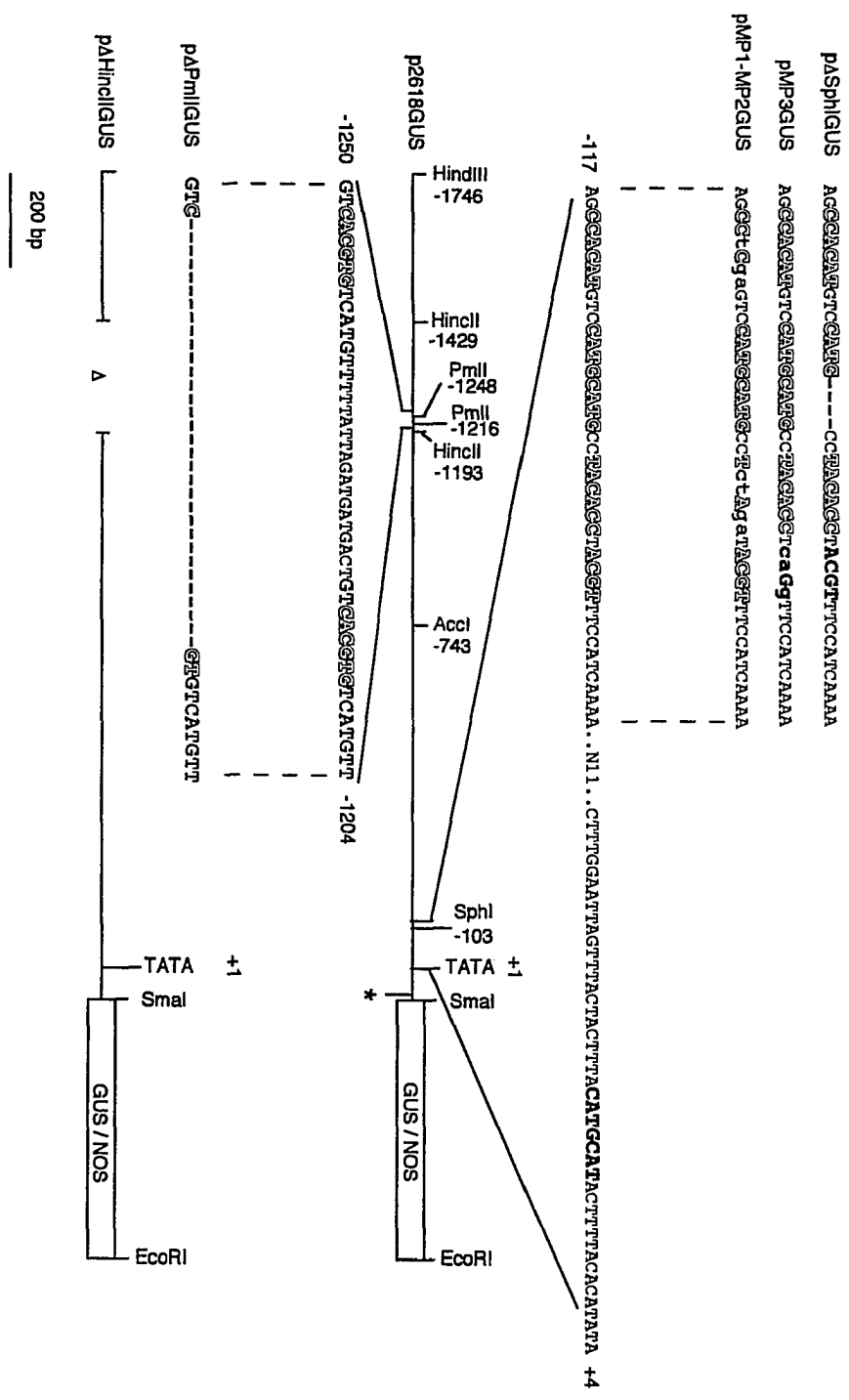


Figure 3. Schematic representation of the six plasmid constructs used in this study. All constructs are transcriptional fusions with the GUS reporter gene and the NOS transcription terminator in pBI121 binary vector. The complete scheme of the control plasmid p2618GUS appears in the center. For the remaining constructs, only the locations where they differ from p2618GUS are shown. Deletions are represented by dashes and base replacements are shown in lowercase letters. The motifs addressed in this study appear in shadow font style. The GUS/NOS fragment is not drawn into scale. Asterisk represents the location of the BglII/BamHI fusion site produced after ligation of the GUS/NOS fragment into the *S/e1* promoter (see experimental procedures for details).



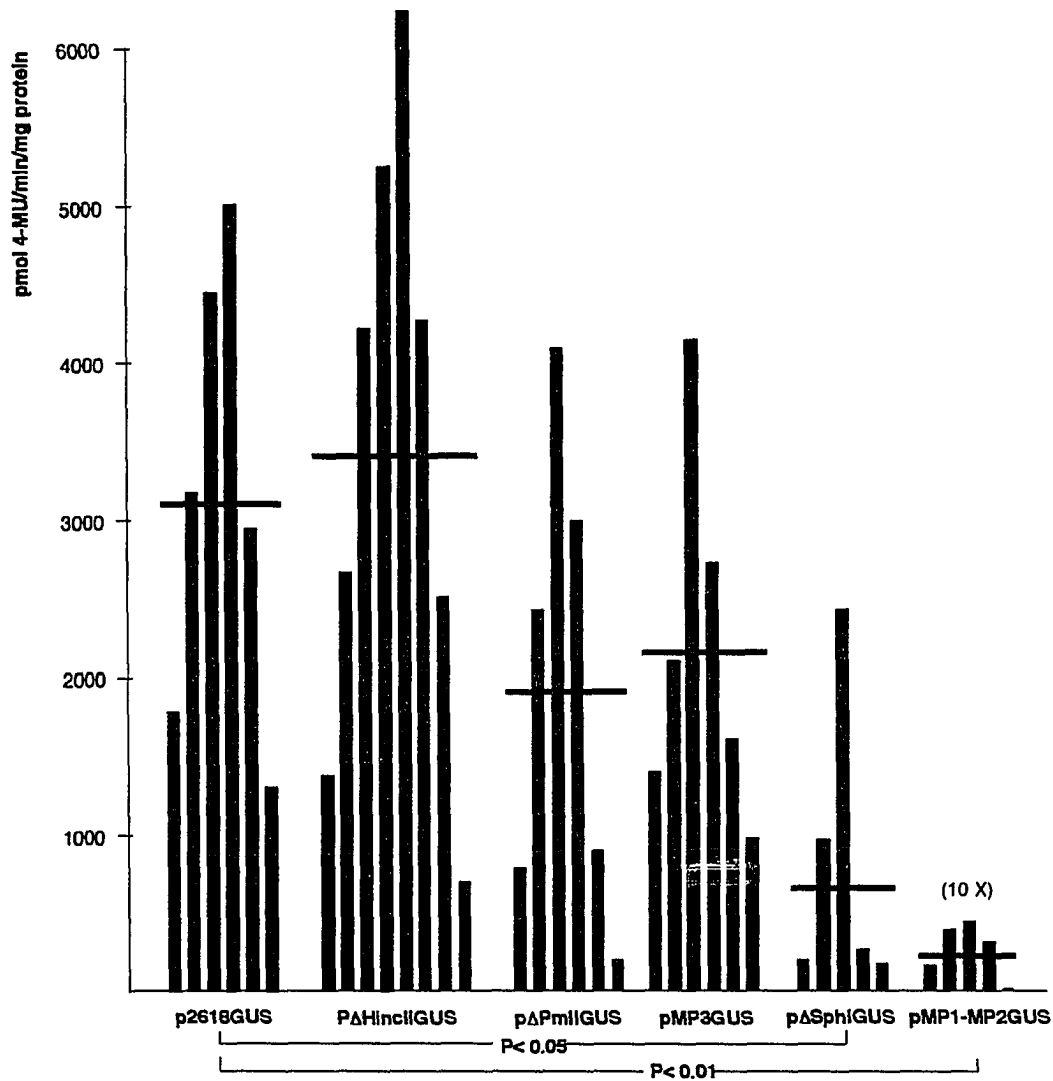


Figure 4. Results of the fluorometric measurements of GUS activity (converted to pmoles of 4-MU/minute/ μ g of total protein) in tobacco mature seeds for six different plasmid constructs. Each bar represents the value of an individual plant. Horizontal bars denote the average value for each plasmid construct. Fluorometric readings of the negative controls (plants not carrying any of the plasmids) were negligible and are not shown in this figure. Due to their low levels of GUS activity, values for plants carrying construct pMP1-MP2GUS were increase by ten times (10 X) in order to fit in the same graphic. The data were submitted to t-test statistical analyses and the two treatments which differ from the positive control (p2618GUS) are connected on the bottom of the figure with their respective significance levels.

Figure 5. Results of the search for CCACAT sequences in other seed-expressed genes containing the SphI box or RY repeat. Panel **A** shows all the genes belonging to Class 1 (genes that contain an identical CCACAT sequence). These genes are *Sle1* and *Sle4* from soybean (Calvo et al., submitted); the *Lea* genes *Emp-1* from rice (Litts et al., 1992) and *Rab-17* from maize (Vilardell et al., 1990); the *Vicia faba* non-storage protein (*USP*; Fiedler et al., 1993); the *Vicia faba* *LeB1* (pseudogene; Heim et al., 1989); and the maize *pML1* (pseudogene; Wandelt and Feix, 1989) storage protein genes. Panel **B** shows examples of Class 2 genes (contain a similar CCACAT motif, with one mismatch allowed) where the majority of genes analyzed fell, and include the maize *22Z-4* (Schmidt et al., 1990), the *Phaseolus vulgaris* phytohemagglutinin (*DLEC1*; Hoffman & Donaldson, 1985), and the *Vicia faba* *LeB4* (Bäumlein et al., 1992) genes as examples. Most genes belonging to Class 2 share a common CCAC within the CCACAT-like sequence. Panel **C** shows two of the three genes which were found to contain a CCACAT-like sequence near the SphI box (Class 3 genes). The examples shown are the soybean *Gy4* (Scallan et al., 1990) and the maize *Opaque-2* gene (*O2*; Gallusci et al., 1994). Notice that several genes belonging to the three different classes have an ACGT core (shown in bold letters) near the SphI box which is also present in the *Sle1* and *Sle4* promoters. For the *O2*, *DLEC1*, and *USP* genes two different sequence intervals are shown for the same gene. For the *USP* gene, the second SphI box (-118 to -154) was the one found to have a negative effect in gene expression (Fiedler et al., 1993). The numbers at the beginning and at the end of each sequence represent the position relative to the TATA box.

A Class 1

pML1 -243 GTCACATGTGTAAAGGTGAAGCGATCATGCATGTCATT-CCACATAGATTAAAAAAA-TCCTATAT -179

USP -300 GCAATGCTGCATGGCATGGCATATTACACCAAAC -268
-154 TTTGAGGTGCATG-CATGGATGCCCTGTGGAAAGTT -118

LeB1 -78 TAACAAATACACCCACATAT.N48.TA-----AATGTCACACACGTTCTGTCACACGTGTT -25

Sle1 -130 CTCTTTCGCCTAAAGCCACATGTC--CATGCATGCC---TACACCTACGTTTCCATCAAATCTAG -70

Sle4 -117 CTCCCTCACCAAAGCCACCTGTC--CATGCATGCC---TACACCTACGTTTCCATCAAATCCAG -60

Emp-1 -334 CCACATGTCCACATGCATGCACGGTGATCGACATCCCGATCGACGTGTCGACA -281

Rab-17 -99 CCCTCACTGGCCGCCCAATCCACTCATGCATGCCACGTACACCCCTGCCGTGGCGCGCCAGATC -34

B Class 2

LeB4 -85 GGCCACAAGC.N48.TAGCCATGCATGCTGAAGAATGTCACACACGTTCTGTCACACGTGTT -25

22Z-4 -233 GTCACATGTGTAAAGGTGAAGAGATCATGCATGTCATT-CCACGTAGATGAAAA-GAATTCCTATAT -169

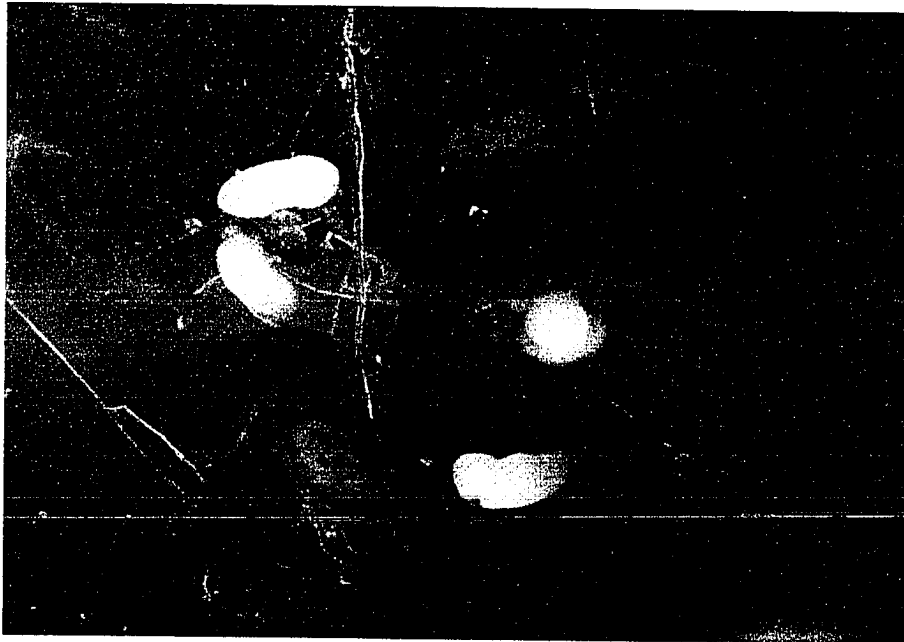
DLEC1 -153 ACTGGCCCTGCTCCATGCATGCATGCACGCTGCCACCACAGCTTT -109

C Class 3

Gy4 -141 ATTGCAAATATAAACGAAGCATGCATGCTTATGTATGTATGAGGTGTAACAAAATTGG -88

O2 -160 GCATGCATGGGTACACGCTTGGCTTATTGCTAACACGTGTGGTTA -116
-106 CACACATGCATGTAGCGTCACGATAAAACCCATGGTACGTAACTTGGT -59

Figure 6. Histochemical GUS staining. The picture shows sectioned dry seeds or dissected intact endosperm and embryos from the control transgenic tobacco plants segregating for the 2618GUS chimeric gene. Note the blue staining in both the seed embryo (e) and endosperm (en). Embryos and endosperm derived from seeds which do not contain the 2618GUS gene fusion are white.



CHAPTER 3: GENERAL CONCLUSIONS

One of the major questions concerning seed development in higher plants is how the different temporal and spatial patterns of gene expression are established in the seed. The understanding of the molecular events that regulate these processes will not only enhance our understanding of the basic aspects of seed development, but may ultimately facilitate our efforts towards the modification of seed composition and properties.

Lea proteins are a set of polypeptides that accumulate during late embryo development and are presumably involved in the protection of embryo cells from the harm of desiccation (Dure, 1993a). This dissertation characterizes an entire *S/e* gene family of Group 1 Lea genes, as the framework to study the regulation of Lea gene expression during soybean seed development and identifies essential sequences involved on the control of the *S/e1* promoter function.

Chapter 1 comprises the cloning, mapping and analyses of expression of the *S/e* gene family in soybean. We initially investigated the structure and organization of the *S/e* gene family in the soybean genome. The five Group 1 Lea genes (*S/e1-5*) could be divided into two classes based on sequence identity. *S/e1-4* were placed on the soybean genetic map and were shown to map to 4 different linkage groups. These findings strongly support the hypothesis of the tetraploid origin of the soybean genome (Grandbastien et al., 1986; Lee and Verma, 1984; Nielsen et al., 1989). *S/e1*, *S/e2*, *S/e3*, and *S/e5* encode polypeptides differing primarily by the presence of a repeated 20-amino acid motif. It remains to be shown whether these polypeptides are functionally equivalent.

Examination of the expression of the *S/e* genes revealed that *S/e* expression was confined to embryo tissues and *S/e* mRNA accumulated at similar

levels in both the embryo axis and in the cotyledons. Members of the same class have similar temporal pattern of expression. Thus, *S/e1* and *S/e5* were shown to be expressed in developing embryos weeks earlier than *S/e2* and *S/e3*. *S/e4* was shown to be a pseudogene which does not contribute to the *S/e* mRNA steady-state levels observed in the seeds. Maximal levels of mRNA for all functional *S/e* genes accumulated before significant desiccation had occurred in the seeds, and declined rapidly upon seed imbibition. Desiccation of seeds did not up-regulate *S/e* expression. Together these findings suggest that seed desiccation does not induce *S/e* expression in the seeds, setting the *S/e* gene family apart from other *Lea* genes in soybean (Hsing et al., 1992) as well as from all the *S/e* homologs in other plant species (Berge et al., 1989; Hughes and Galau, 1991; Espelund et al., 1992; Butler and Cuming, 1993; Quatrano et al., 1993).

Exogenous ABA was not required to induce *S/e* expression in cultured embryos and only moderately increased *S/e* mRNA levels. Neither ABA nor desiccation induced expression of the *S/e* genes in vegetative tissues. These results demonstrate that the *S/e* genes are less ABA dependent than its counterparts in wheat (Williamson and Quatrano, 1985), barley (Espelund et al., 1992), and maize (Williams and Tsang, 1991; Butler and Cuming, 1993).

In Chapter 2 we investigated the nucleotide sequence of the promoter region for *S/e1-4*. We found that the *S/e1* and *S/e4* promoters each have two conserved SphI box motifs (CATGCATG) in their proximal promoter regions, one of them flanked by two CCACAT inverted repeats. Interestingly, two SphI boxes, but not the CCACAT repeat, are found at similar positions in the promoter of the soybean storage protein genes where they act as major seed-specific enhancers (Chamberland et al., 1992; Lelivre et al., 1992). Two hexameric G-box core motifs (CACGTG) found in the promoter of most *Lea* genes (Quatrano et al., 1993) are

also present at the distal promoter region of *S/e1*. *S/e2* and *S/e3* promoters contain only variants (CAATGCATG) of the SphI box and are devoid of the CCACAT repeats, but do contain the G-box core at the proximal promoter region.

Transcription fusion of a 1823 bp *S/e1* promoter fragment to the GUS coding region was able to drive GUS gene expression in both the embryo and the endosperm of transgenic tobacco seeds. This observation contrasts with the embryo-specific expression conferred by the promoter of the *S/e* homolog from wheat in tobacco seeds (Marcotte et al., 1989), and substantiates the possible existence of somewhat different cis-acting requirements in the promoter of homologous *Lea* genes across different species.

Several mutant versions of the *S/e1* promoter were also analyzed. The results show that neither the two CACGTG hexameric core motifs located at positions -1312 and -1344 bp nor a third ACGT-containing motif at -89 bp constitute a major site for *S/e1* promoter function in transgenic tobacco seeds. Rather, the SphI box at position -106 bp and its flanking CCACAT repeats play a major positive role in this response. This represents the first report of the involvement of these sequence motifs in *Lea* gene expression. It remains to be determined how the SphI box acts to bring about the positive effect in the promoter of soybean genes with very distinct temporal patterns of expression such as the glycinin (Lelivre et al., 1992) and β -conglycinin (Chamberland et al., 1992) storage proteins versus the *S/e1* *Lea* gene.

Since *S/e2* and 3 appear to have a somewhat different proximal promoter, it would be valuable to determine the sequence requirements for their promoter function and whether the SphI box variant CAATGCATG is a major site for promoter function or if its role is diminished by the nearby G-box core motif. Moreover, nuclear protein extract binding assays on the different *S/e* promoters may uncover

some of the proteins involved in *Lea* gene expression during soybean seed development. The abundance of embryogenic material in the soybean seed should facilitate this type of experiment.

Taken together, the results from Chapter 1 and 2 suggest that an ABA-dependent transduction pathway, acting upon the G-box core motifs or on the proximal ACGT core, does not appear to be the main regulatory pathway controlling *S/e1* gene expression during soybean late embryo development, as has been proposed for the *S/e1* homolog in wheat (Marcotte et al., 1990; Guiltinan et al., 1990; McCarty et al., 1991; Quatrano et al., 1993). Therefore, other yet uncharacterized pathway(s) acting upon the SphI box and its flanking CCACAT sequences are likely to play a major role in *S/e1* promoter function.

Literature Cited

- Chamberland S., Daigle N., Bernier R. (1992) The legumin boxes and the 3' part of a soybean β -conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. *Plant Mol. Biol.* 19: 937-949.
- Guiltinan M.J., Marcotte Jr.W.R., Quatrano R.S. (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* 250: 267-271.
- Lelievre J.-M., Oliveira L.O., Nielsen N.C. (1992) 5'-CATGCAT-3' elements modulate the expression of glycinin genes. *Plant Physiol.* 98: 387-391.
- McCarty D.R., Hattori T., Carson C.B., Vasil V., Lazar M., Vasil I.K. (1991) The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66: 895-905.
- Quatrano R.S., Marcotte Jr.W.R., Guiltinan M. (1993) Regulation of gene expression by abscisic acid In: Ed. Verma DPS *Control of Plant Gene Expression*, 69-90. CRC Press, Boca Raton, FL.

ACKNOWLEDGMENTS

I Would like to thank Dr. Randy C. Shoemaker for his guidance during my years at ISU. I really appreciate the time, patience, and effort he has invested in my training.

I also would like to express my gratitude to all the members of the Shoemaker's lab during the past five years. I certainly have learned a great deal from all of them.